UNIVERSITY OF OSLO
Deptartement of Linguistics and
Scandinavian Studies

Automatic Anaphora Resolution for Norwegian (ARN)

Thesis submitted for the degree of
Candidata Philologiæ

Gordana Ilić Holen

May 24, 2006

Takk

Jeg vil først og fremst takke min veileder Janne Bondi Johannessen for hennes meget gode veiledning. Jeg satt veldig pris på hennes klare tilbakemeldinger – jeg fikk ros når oppgaven gikk riktig vei, og jeg fikk klar og konstruktiv kritikk når den ikke gjorde det. Janne ga meg mange anledninger til å presentere oppgaven min, i begynnelsen ved å invitere meg på seminarer, og mot slutten ved å oppfordre meg til å delta på konferanser, noe både oppgaven min og jeg hadde stort utbytte av. Jeg takker også resten av Tekstlaboratorietlaget og spesielt Anders Nøklestad for hjelp til å komme meg ut av korpusforviklinger.

En halvtimes programmeringsveiledning fra Paul Meurer tidlig i 2003 ga meg nok å tygge på og arbeide med i mange måneder.

Det håndannoterte korpuset jeg fikk fra BREDT-prosjektet ved Universitetet i Bergen var en uvurderlig ressurs for trening og evaluering av mitt prosjekt.

Og så til venner: Takk til Øyvind Lie for alle de fine kaffepausene gjennom våre mange år på Blindern. Takk til Andra Björk og Cecilie for mange gode faglige diskusjoner og for å ha vært mine venner gjennom tykt og tynt.

Takk til min familie i Serbia, Nederland og Norge, og spesielt til min mann Øistein for kjærlighet og støtte. Og takk til lille Eir, fordi hun finnes.

Hvala

Na prvom mestu želim da se zahvalim mojoj mentorki Jane Bondi Johanesen. Ona je umela da pohvali kada bi rad išao kako treba i da pruži jasnu i konstruktivnu kritiku kada ne bi. Jane mi je pružila mnogo mogućnosti da prezentujem svoj rad, u početku pozivajući me na seminare, a kasnije podstičući me da se prijavim na konferencije, od čega smo i ja i moj rad imali mnogo koristi. Takodje sam zahvalna ostatku tima Tekst laboratorije, a posebno Anešu Noklestadu, na pomoći u radu sa korpusom.

Polučasovni razgovor sa Paulom Mojrerom početkom 2003. mi je pružio dovoljno inspiracije za više meseci programiranja.

Manuelno anotirani korpus koji sam dobila od BREDT projekta sa Univerziteta u Bergenu je bio nezamenljiva pomoć u treniranju i evaluaciji mog projekta.

Hvala i prijateljima: Ejvindu Liju na svim pauzama za kafu za naših mnogo godina na Blindernu, Andri Bjork i Sesilie na dobrim stručnim diskusijama i zato što su mi bile prijateljice u dobrim i lošim vremenima.

Hvala mojoj porodici u Srbiji, Norveškoj i Holandiji, a posebno mom mužu Ejstajnu, na ljubavi i podršci. I hvala maloj Eir, zato što postoji.

HEAD KNIGHT: Then, when you have found the shrubbery, you must cut down the mightiest tree in the forest... with... a herring!
ARTHUR: We shall do no such thing!
HEAD KNIGHT: Oh, please!
ARTHUR: Cut down a tree with a herring? It can't be done.
KNIGHTS: Aaaaugh! Aaaugh!
HEAD KNIGHT: Don't say that word.
ARTHUR: What word?
HEAD KNIGHT: I cannot tell, suffice to say is one of the words the Knights of Nee cannot hear.
ARTHUR: How can we not say the word if you don't tell us what it is?
KNIGHTS: Aaaaugh! Aaaugh!
ARTHUR: What, 'is'?
HEAD KNIGHT: No, not 'is' – we couldn't get vary far in life not saying 'is'.
BEDEMIR: My liege, it's Sir Robin!
(. . .)
ARTHUR: Oh, Robin!
ROBIN: My liege! It's good to see you!
KNIGHTS: Aaaaugh!
HEAD KNIGHT: He said the word!
ARTHUR: Surely you've not given up your quest for the Holy Grail?
MINSTREL (singing): He is sneaking away and buggering up–
ROBIN: Shut up! No, no no– far from it.
HEAD KNIGHT: He said the word again!
ROBIN: I was looking for it.
KNIGHTS: Aaaaugh!
ROBIN: Uh, here, here in this forest.
ARTHUR: No, it is far from–
KNIGHTS: Aaaaugh!
HEAD KNIGHT: Aaaaugh! Stop saying the word!
ARTHUR: Oh, stop it!
KNIGHTS: Aaaaugh!
HEAD KNIGHT: Oh! He said it again!
ARTHUR: Patsy!
HEAD KNIGHT: Aaugh! I said it! I said it! Ooh! I said it again!
KNIGHTS: Aaaaugh!

The Knights of Nee on the subject of **'It'**
from *Monty Python and the Holy Grail*

Contents

Chapter 1

Introduction

1.1 On ARN

In this thesis **ARN – an Automatic Anaphora Resolution System for Norwegian**[1] is presented. I have developed ARN as a rule-based anaphora resolution system, on the basis of two existing systems for English: MARS and RAP[2]. However it turns out that anaphora resolution systems cannot freely be transported across individual languages. During the course of work with ARN, I have discovered that many of the rules that were successfully used for English cannot be applied for Norwegian, due to not very known information structure differences between the two languages.

ARN has been designed to resolve the third person pronoun with the exception of pronoun *det* 'it (neut.)'. On this it has achieved an accuracy of 70.5%.

1.2 The task of anaphora resolution

Anaphora are words that specify a real-world entity by refering through another textual item, *antecedent* (Hirst 1981). The term itself comes from the two Greek words, $\alpha\nu\alpha$ meaning *back, upstream* and $\phi o\rho\alpha$ - *the act of carrying*, so the term itself could be understood as *the act of carrying back upstream* (Denber 1998).

I will introduce the challenges of the task of automatic anaphora resolution through an example. This can also be an illustration of the basic terminology related to the task.

[1] By 'Norwegian' I mean here and in the rest of the thesis the Bokmål variant of the Norwegian language.

[2] Mitkov (2001) and Lappin and Leass (1994) respectively.

(1) But Mr Prodi will have to wait until after the election of a new Italian head of state in mid-May before he_1 can actually begin to appoint ministers from among his centre left coalition supporters and begin to run the country.

Despite this, Mr Berlusconi has remained defiant.

Speaking to supporters in the northern city of Trieste, he_2 said he_3 had no intention of making any formal telephone call to Mr Prodi conceding defeat, as **he**$_4$ believes the new centre left coalition will quickly become unglued, reports the BBC's David Willey in Rome.

There are several anaphora in this text, but let us concentrate on **he**$_4$. *Resolving* this anaphor, i.e. finding its antecedent would go automatically for a human reader, without thinking of it. Only if she has her mind somewhere else while reading she might stop for a moment and think: "Who, Prodi or Berlusconi?". For a machine the possible *antecedent candidates* are, apart from (for us the obvious) Prodi and Berlusconi, also, at least: *the election of a new Italian head of state in mid-May, the election, a new Italian head of state, state, mid-May,* he_1*, ministers, his centre left coalition supporters, centre, left, coalition, supporters, the country, this, speaking to supporters, speaking, supporters, the northern city of Trieste,* he_2*,* he_3*, intention of making any formal telephone call to Mr Prodi conceding defeat, intention, making any formal telephone call to Mr Prodi, formal telephone call to Mr Prodi, telephone call, telephone, call, conceding defeat, conceding* and *defeat.* So for a machine the task of anaphora resolution is not trivial at all.

In natural languages, anaphora are essential part of the cohesional forces that keep the discourse together. This makes anaphora resolution highly important for numerous natural language processing (NLP) applications, such as natural language interfaces, automatic text abstracting, information extraction and machine translation.

1.3 Why a system for Norwegian?

ARN is an anaphora resolution system designed especially for Norwegian. What makes Norwegian so special that it needs its own anaphora resolution system? Most of the AR systems are made for English, including the systems ARN is directly inspired by: MARS (Mitkov 2001) and RAP (Lappin and Leass 1994). I will present these systems at greater length in chapter 3, but I will here present some of their features that could provide an answer to this question. Both systems are rule-based, containing sets of rules which award points to antecedent candidates during the resolution of an anaphor. The candidate with the most points is ultimately proposed as the antecedent of the anaphor in question. Most of the rules

the systems have in common are based on the following super-rule:

> Prefer candidates that are subjects to candidates that are direct objects, prefer direct objects to indirect objects and prefer indirect objects to other constituents such as adverbial or prepositional phrases.

As many as four out of seven rules applied by RAP are based on this super rule, and RAP is with its 86% of correctly resolved anaphora considered an exceptionally successful AR system. In addition, the rules in both RAP and MARS can award candidates different point sums, and both systems choose the highest sums for candidates that are subjects. So, to the point: this hierarchy does not work for Norwegian. Giving preference to subject candidates always impairs the system's performance, and so does the penalizing of prepositional phrase candidates. Awarding object candidates and penalizing the candidates that are the part of an adverbial phrase is not very helpful either, as introducing these rules has minimal impact on the system's results.

This is a surprising find, considering the fact that Norwegian and English are closely related languages, and the fact that MARS has been successfully applied to such different languages as Polish and Arabic. With some minor modifications that did not involve the rules based on the aforementioned super-rule, MARS achieved quite good results on these languages, urging Mitkov to propose that rules applied in MARS can be regarded as "multilingual".

I have come to the conclusion that this is due to the differences in information structure between Norwegian and English. Although there is a preference in both languages not to convey the new information by subject, this preference is much stronger in Norwegian. This tendency leads to a much higher number of constructions with the expletive subject 'it' (Norw. 'det') in Norwegian than in English, which not only leaves a considerable number of subjects quite unsuitable for antecedent candidates, but also influences the salience of direct and indirect objects and adverbial and prepositional phrases.

I will discuss this topic at length in chapter 5, but we could already now conclude that there are good indications that Norwegian does need an anaphora resolution system of its own.

1.4 An overview of the thesis

The thesis itself consists of seven chapters. In chapter 1 I have introduced ARN. The chapter 2 will present the anaphora and their possible antecedents, i.e. it will contain a short overview of Norwegian third person pronouns and Norwegian nouns with their gender and number features.

The chapter 3 will describe the main anaphora resolution systems ARN has been inspired by. The chapters 4 and 5 describe ARN itself: the chapter 4 gives its architecture - the organisation of data and implementation of the system, while the chapter 5 takes a closer look at the resolution rules implemented by ARN and their impact on the training part of the corpus. In this chapter I have discussed many of the problems that English-based rules create for a Norwegian anaphora resolution system. The chapter 6 compares the results ARN achieved on the test corpus with baseline systems and the anaphora resolution systems described in chapter 3. And finally, the chapter 7 takes a look back at what ARN have achieved and a look ahead at what ARN could be developed into.

Chapter 2

Antecedents and candidates

In this chapter we will have a look at the scope of ARN, the anaphora it attempts to resolve – the Norwegian third person pronoun and the possible resolution candidates – Norwegian nouns with their number and gender features.

2.1 The scope of ARN

ARN has been designed to be able to handle resolution of all personal pronouns in Norwegian, but its prime aim is to resolve the third person pronouns with the exception of the pronoun *det* 'it (neut.)'. In contrast to the third person pronouns, the first and the second person pronouns have their antecedent either outside the text, or they appear in dialogues. This posed logistic problems as the corpus used for training of ARN did not have tags that indicate dialogues. Resolution of the first and second person pronouns could be included in later versions of ARN.

The problem of excluding the pronoun *det* 'it (neut.)' from ARN's scope is possibly more complex. In English, the resolution of the anaphor *it* is still one of the larger unsolved problems in anaphora resolution. The case of the Norwegian *det* 'it (neut.)' is similar, though at some points *det* is even more complicated than the English *it*. Here is an overview of some of the problems one would encounter when resolving this anaphor:

- The anaphor *det* is ambiguous with the definite article *det*, as in the example (2):

  (2) det røde eplet
  the/it red the-apple
  the red apple

  This is a minor problem, as it is possible to distinguish the anaphor from the article by means of syntactic analysis of the noun phrase.

- The anaphor *det* is ambiguous with the expletive subject *det*. This poses a more serious problem as the contexts in which *det* has no reference and in which *det* is an anaphor do not differ much. Consider the two sentences in examples (3) and (4):

  (3) Det kom en mann nedover veien.
  it came a man down the-road
  A man came down the road.

  (4) Det kom nedover veien.
  it came down the-road
  It came down the road.

  In example (3), *det* is a formal, expletive subject. As a mere place holder it is devoid of meaning and does not refer to anything. On the other hand, *det* in the example (4) can refer to any entity of singular number and neutral gender, such as *barn* 'child (neut.)' or kjøretøy 'vehicle (neut.)'.

- One of the most important problems in connection with *det* is that it can have non-nominal antecedents:

  (5) John sparka Tom. Det var ikke særlig kult.
  john kicked tom it was not particularly cool
  John kicked Tom. That wasn't very cool.

  In the example (5), *det* refers to the whole situation of John kicking Tom. When *det*'s antecedent is a situation, it can textually spread over several sentences, even whole paragraphs. Expanding the search for antecedent candidates from nominal phrases in the vicinity of the anaphor to whole sentences and even paragraphs is still one of the unsolved problems in anaphora resolution in general, and certainly too big an enterprise to be fit into the time-frame of this thesis.

2.2 The third person pronouns

Apart from having different forms for different genders and numbers, the third person pronouns in Norwegian also differ for human and non-human entities. The main rule is that pronouns *han* and *hun* refer to humans, while *den* and *det* refer to other living beings, not-living objects and abstracts:

(6) Har du sett broren min? Han står der borte.
have you seen the-brother mine(m.sg.) he stand there away
Have you seen my brother? He is standing over there.

(7) Har du sett sykkelen min? Den står der borte.
have you seen the-bicycle mine(m.sg.) it(m.) stand there away
Have you seen my bicycle? It is standing over there

	Singular					Plural
	Human			Non-Human		
	m.	f.	Polite (m. & f.)	m. & f.	*n.*	
Subject form	han	hun	De[1]	den	*det*	de
Object form	han ham	henne	Dem	den	*det*	dem

Table 2.1: The third person pronouns in Norwegian

This, main, reading is presented in the table 2.1. There are, however, exceptions on both sides of the human/non-human division line. *Hun* can be used for referring to ships, while both *han* and *hun* can be used to refer to non-humans, as for example pets and animals in general. *Det* is also used for referring to the nouns of neutral gender that denote humans. Those words are not many, but there are widely used, such as nouns *barn* 'child' and the word 'human' itself (*menneske*). *Den* is possibly even rarer then *det* in denoting humans. The only example I could come over is using *den* to refer to *unge* 'kid'. In an automatic anaphora resolution system it is difficult to account for those exceptions, so the only ones covered are the neuter words that denote humans: the aforementioned *menneske* and *barn* as well as the other compound kinship words that include the word *barn* such as *barnebarn* 'grandchild' and *stebarn* 'stepchild'.

2.3 Nouns

In this section I am going to present the possible antecedent candidates, i.e. nouns. For the resolution of anaphora, it is important if nouns denote a singular entity or not; the gender of nouns is also important, but even more so if they denote humans and their sex.

2.3.1 Gender

Norwegian has two gender systems: one not entirely consequent three-gender system consisting of masculine, feminine and neuter and a two-gender system where feminine and masculine gender have coalesced into the one, common gender.

The nouns normally have one gender that is constant throughout the declension, so it is considered an inherent quality of a noun in contrast to the categories of number and definiteness. In most cases it is not possible to determine gender from the form of the noun, but gender is decisive for the form of the the words determining the noun (e.g. determinatives and adjectives), and for the definite form of the noun.

The three-gender system consists of:

- masculine

  (8) en liten stol
  a(m.) little(m.) chair
  a little chair

  (9) den lille stolen min
  the(m./f.) little(def.) chair mine(m.)
  my little chair

- feminine

  (10) ei lita hylle
  a(f.) little(f.) shelf
  a little shelf

  (11) den lille hylla mi
  the(m./f.) little(def.) shelf mine(f.)
  my little shelf

- neuter

  (12) et lite bord
  a(n.) little(n.) table
  a little table

(13) det lille bordet mitt
the(n.) little(def.) table mine(n.)
my little table

In the two-gender system, there is a common and neuter gender. The forms of the common gender are identical to those of the masculine gender in the three-gender system:

- common gender

(14) en liten stol
a(m./c.) little(m./c.) chair
a little chair

(15) den lille stolen min
the(m./c.) little(def.) chair mine(m./c.)
my little chair

(16) en liten hylle
a(m./c.) little(m./c.) shelf
a little shelf

(17) den lille hyllen min
the(m./c.) little(def.) shelf mine(m./c.)
my little shelf

- neuter

(18) et lite bord
a(n.) little(n.) table
a little table

(19) det lille bordet mitt
the(n.) little(def.) table mine(n.)
my little table

It is not unusual to combine the two systems, and the following combination is relatively widespread today (Faarlund et al. 1997, p. 152):

(20) en bok
a(m./c.) book
a book

(21) min bok
my(m./c.) book
my book

(22) boken *or* boka
the-book(m./c.) *the-book(f.)*
the book

(23) boken min *or* boka mi
the-book(m./c.) mine(m./c.) *the-book(f.) mine(f.)*
my book

Another problematic issue is the conflict between (the natural category of) sex and (the grammatical category of) gender. This conflict affects both the two- and the three-gender system. Many of the nouns referring to persons of both sexes are of masculine gender, including such huge noun groups as nouns referring to:

- Someone living in or originating from some place, like *haugesunder* 'person from Haugesund', *tysker* 'German', *russer* 'Russian'
- Persons engaged in some kind of activity, persistent or temporary, such as *drømmer* 'dreamer', *eventyrer* 'adventurer', *språkbruker* 'language user', *vegetarianer* 'vegetarian', *tyv* 'thief', *angriper* 'attacker', *beboer* 'inhabitant', *kinogjenger* 'cinema-goer', *nabo* 'neighbour', *elev* 'pupil'
- Professions: *adjunkt* 'teacher', *advokat* 'solicitor', *baker* 'baker', *biolog* 'biologist', *dommer* 'judge', *forsvarsminister* 'Minister of Defence', *førstebetjent* 'police officer', *kokk* 'cook', *lærling* 'apprentice', *sjef* 'boss'

Words with feminine gender that denote male persons do exist but are few in numbers. The only two examples I have been able to find are *vakt* 'guard' and *pyse* 'sissy'.

The conflict between sex and gender can pose problems for anaphora resolution. Consider the following example:

(24) Førstebetjenten kikket på Anna og fortsatte å lese i
the-officer(m.sg.) looked on Anna and continued to read in
boka. Hun virket ikke særlig bekymret.
the-book(f.sg). she seemed not particularly worried

The officer looked at Anna and continued reading the book. She didn't seem particularly worried.

In example 24, the correct antecedent of the anaphor *hun* 'she' is *førstebetjenten* 'the officer', which is of the masculine gender. If we ignore the sex/gender conflict we may discard the correct candidate at a very early stage of resolution. To account for this, a new classification of nouns and pronouns which merges sex and gender has been made for ARN. This classification will be presented in section 5.3.1.

2.3.2 Number and definiteness

The Norwegian nouns inflect for definiteness and number, both of which are dependent on the noun's gender. With respect to anaphora resolution, the category of number is more important since there are different pronouns that refer to singular and plural entities.

The Norwegian language has two numbers, singular and plural. Nouns of feminine, masculine or common gender normally do have a morphological expression of number. However, a substantial group of nouns of neuter gender has the same form in singular and plural, which leads to ambiguity that in itself can result in weaker results of an anaphora resolution system.

The nouns referred to by the plural pronoun *de* 'they' are usually in their plural grammatical number, so the discrepancy between the natural and the grammatical category of number is not as substantial as the case is with the categories of gender and sex. There are however some examples of it, such as words *regjering* 'government' and *folk* 'people'.

Lastly, the definiteness of a noun has some influence on the anaphora resolution as definite nouns tend to be more salient. I will discuss how big this influence is when I present the resolution rule that penalizes the antecedent candidates with indefinite form in chapter 5.

Chapter 3

The theoretical background

In this chapter I will first give a short presentation of the field and then give a more detailed description of the two anaphora resolution systems ARN is inspired by: Mitkov's Original Approach with its later implementation MARS (Mitkov 1998; Mitkov 2002) and RAP (Lappin and Leass 1994). Mitkov's systems and RAP are only presented in this chapter, without comment. When I come to the performance of my system in the Evaluation chapter (ch. 6), I will have a more thorough discussion over which parts of these systems were found useful, what was improved and what was discarded.

3.1 Anaphora resolution systems - a short overview

Anaphora resolution (AR) has been a subject of study for both theoretical and computational linguistics for a long time. Although pronominal anaphora does not constitute more than 15% of coreference annotations in the texts annotated by humans (Baldwin 1995), most of the anaphora resolution systems that exist today deal with pronominal anaphora. The resolution of nominal anaphora such as definite descriptions and indirect anaphora is much less represented in modern systems (Mitkov 2001).

The focus on pronominal anaphora dates back to the artificial intelligence (AI) systems of the 1970's where anaphora resolutioners were implemented as a part of the language modules. The language module of SHRDLU (Winograd 1972) did pronominal resolution by forming a history list consisting of noun phrases mentioned in previous sentences, and then applying relatively simple heuristics based on selectional restrictions and syntactic constraints on pronominalization.

In the 1980's, anaphora resolution was approached as a problem in its own right, so most of the AR systems attempted resolution of both pronominal and nominal anaphora. The systems relied heavily on linguistic knowledge, and tried

to limit the amount of pragmatic (world) knowledge used. Many of the systems were never or only partially implemented, such as Rich and LuperFoy (1988). Some achieved good results, but on very limited corpora, for instance Carbonell and Brown (1988) achieved 87% on a sample of 31 anaphora. The systems were exceptionally computationally costly, and often the most expensive part was the representation of world knowledge. Consider the following example:

(25) The soldiers shot at the women and they fell.
The soldiers shot at the women and they missed.[1]

Although the only way to resolve anaphor *they* is to apply world knowledge, the price seemed too high to pay, so in the 1990's applying this type of knowledge was mostly abandoned. The AR systems of the 1990's towards our time have been greatly influenced by two developments:

- NLP applications such as information extraction, natural language interactions and automatic abstraction started to become widely used, which created a need for a quick and reliable anaphora resolution in real-life environment.

- Quick development and greater availability of NLP tools such as word nets (ontologies), POS-taggers, parsers and corpora.

The need for real-time AR has inspired the design of knowledge-poor systems. Those systems deliberately limit not only pragmatic, but all kinds of knowledge. The systems ARN is inspired by belongs to this category, and they will be presented in greater detail at the remainder of this chapter. The representative AR systems for this line of development are MOA (Mitkov 1998), MARS (Mitkov 2001), the AR system by Kennedy and Boguraev (1996) and, to a certain extent, RAP system by Lappin and Leass (1994). These systems are often based on (or include) a set of rules that award antecedent candidates according to their salience, thus promoting the most salient candidates.

It is, of course, not the case that all of the AR systems have suddenly discarded all knowledge sources. Many of the systems include the rules of Chomsky's (1981) Government and Binding theory such as Stuckardt (2001) and Lappin and Leass (1994). The Centering theory (Grosz et al. 1995) is also used as the theoretical basis for AR systems, such as the approach by Strube and Hahn (1999).

Concerning the availability of new NLP tools, it is the appearance of large corpora that has had the strongest influence on AR. In the AR systems of the 1980's it was not unusual that the authors themselves had to provide both the training and the testing corpora for their systems. This was time-consuming, and the texts

[1] Example from Carter (1987)

had often little in common with the real-life challenges. The SPAR system (Carter 1987) was, for instance, trained and tested on a corpus of very short stories written either especially for SPAR or for other NLP systems.

The availability of corpora induced strong development of statistical approaches to AR. It started with a large-scale pattern-matching (Dagan and Itai 1990), but it is the unsupervised machine learning and probabilistic approaches that are today the area of most interest (Aone and Bennett 1995; Ge, Hale, and Charniak 1998).

And finally, there is the combination of approaches, which in my opinion has the greatest growth potential, although results obtained so far are not very impressive. Orasan et al. (2000) made statistical optimization of factor points for MARS system (Mitkov 2001), but the results improved by less than two percentage points, ending up at 61.55%. Dagan et al. (1995) made a statistical expansion of RAP called RAPSTAT, which used statistically measured lexical preference patterns to re-evaluate candidates suggested by RAP. The initially good performance of RAP was increased by 3 percentage points to 89%.

For the Scandinavian languages a couple of systems have recently been developed. Martin Hassel (2000) made the AR module for the automatic summarizing system SweSum[2]. His system was based on the knowledge-poor algorithm developed by Kari Fraurud (1992) which had earlier been successfully implemented on Swedish texts. In Norway, the BREDT project[3] at the University of Bergen, lead by Christer Johansson, aims to use relatively simple statistical methods to identify referential chains. This project is still in progress.

3.2 The theoretical basis

The two main anaphora resolution systems that inspired ARN are Mitkov's Original Approach (MOA) (Mitkov 1998) with its further development MARS (Mitkov 2001), and Lappin and Leass' RAP (1994). These approaches can be classified as rule-based knowledge-poor systems: the anaphora resolution modules contain a set of rules that award or penalize antecedent candidates. Mitkov calls the rules *antecedent indicators*, while Leass and Lappin refer to them as *salience factors*.

3.2.1 Mitkov's Original Approach (MOA) and Mitkov's Anaphora Resolution System (MARS)

The making of Mitkov's Original Approach (MOA) AR system was inspired by the need for anaphora resolution in real-life knowledge-poor environments (Mitkov

[2]http://swesum.nada.kth.se/index-eng.html

[3]http://bredt.uib.no/index_eng.html. The BREDT has provided the corpus manually tagged for reference that was crucial for ARN's training and evaluation.

2002). In order to maximally cut down the linguistic knowledge and thereby computational costs, it was not provided with a parser, only with a part-of-speech (POS) tagger. This system was later modified and reimplemented, and this new version is known as Mitkov's Anaphora Resolution System (MARS) (Mitkov 2002). The main difference between the two systems is that MARS is fully automated; it is also provided with a parser, and some new antecedent indicators. Both systems were trained and tested on texts that belong to the genre of technical manuals.

The operation of the systems can be outlined in six steps:

1. The text is POS tagged (MAO) or parsed (MARS).
2. NPs are extracted by an NP-extractor.
3. NPs that precede the anaphor are located. MOA locates the NPs in a two-sentence scope, while MARS looks for antecedents in all of the text that precedes the anaphor.
4. Gender and number constraints are applied.
5. Antecedent indicators are applied to the antecedent candidates that agree in gender and number. The scores (2, 1, 0 or -1) are assigned.
6. The NP with the highest score is proposed as antecedent. In the case of a tie:
   (a) the antecedent candidate with the higher score for immediate reference (3.2.1, page 19) is chosen.
   (b) if this does not resolve the tie or if immediate reference has not been identified, the priority is given to the candidate with the highest collocation match score (3.2.1, page 18) .
   (c) if still undecided, the most recent of the remaining candidates is selected.

Antecedent indicators

The anaphora resolution module is based on a list of rules called *antecedent indicators*. Each antecedent indicator assigns a score of -1, 0, 1, or 2 to each candidate. The candidate with the highest overall score is proposed as the antecedent. For an overview, a list of all the indicators will be given, and then each will be examined more closely.

- *Boosting antecedent indicators.* The boosting indicators promote antecedent candidates by giving them positive points (+1 or +2). In the original version of the system (MOA), Mitkov lists eight indicators:

1. First noun phrases/ Givenness
2. Indicating verbs
3. Lexical reiteration
4. Section heading preference
5. Collocation match
6. Immediate reference
7. Sequential instructions
8. Term preference

In MARS three more indicators are added:

9. Boost pronoun
10. Syntactic parallelism
11. Frequent candidates

- *Impeding antecedent indicators.* The impeding indicators are used to discourage candidates by giving them negative points:
  1. Indefiniteness
  2. Prepositional NPs (NPs that are part of prepositional phrases)
- *The antecedent indicator that function both boosting and impeding*
  1. Referential distance

First noun phrases/ Givenness. In MOA, the first NPs of the sentence are awarded +1 points. The argumentation for this is twofold:

1. Subject salience
2. Theme salience

As MOA operates on texts pre-processed by a POS-tagger, it does not have the syntactic information on subject, but it assumes that the first NP is the subject, English being a SVO language. If the first NP is considered a subject, the subject preference as proposed by the Centering theory (Grosz et al. 1995) can be applied. In addition to this, Mitkov uses a heuristics that the first NP in a non-imperative sentence is the known (given) information (*theme*), while *rheme* which introduces new information on theme, is introduced later. The first NP is then the most

salient one, since in a coherent text theme usually appears first, and thus forms a coreferential link with the preceding text (Firbas 1992).

In contrast to MOA, MARS does have a parser, and thus the information on syntactic roles. The antecedent indicator is here renamed **Obliqueness**. Following the Centering theory where grammatical function is used as discourse salience indicator, MARS awards subject NPs a score of +2, direct object a score of +1, indirect object gets no bonus and the NPs for which the parser cannot identify the function are penalized with -1.

Indicating verbs. If a verb is a member of the given closed class Verb_set, the first NP following it is considered the preferred antecedent and awarded a score of +1.

> Verb_set = {*discuss, present, illustrate, identify, summarize, examine, describe, define, show, check, develop, review, report, outline, consider, investigate, explore, assess, analyse, synthesize, study, survey, deal, cover*}

Empirical evidence suggests that the verbs in Verb_set are good indicators that NPs that immediately follow them carry more salience.

Lexical reiteration. The score of +2 is assigned to NPs repeated twice or more times in the paragraph, the score +1 to the NPs repeated once, and NPs that are not repeated are given 0 points. Lexically reiterated items include repeated synonymous noun phrases as well as sequences of noun phrases with the same head, e.g. “toner bottle”, “bottle of toner”, “the bottle”.

The weakness of this indicator is that it does not capture the fact that the salience of a noun also increases if pronouns refering to it have been repeated. This was corrected in the MARS version, where also anaphora previously resolved to a noun have been taken into account.

Section heading preference. A score of +1 is awarded to NPs that also appear in the heading of the section the analyzed sentence appears in. This score is awarded in addition to the score obtained through the lexical reiteration indicator.

Collocation match. This indicator awards candidates that immediately precede or follow a verb that is identical or morphologically related to a verb that immediately precedes or follows the anaphor in the same paragraph. This indicator is restricted to the following patterns:

> <NP/Pron, Verb>, <Verb, NP/ Pron>

and

<NP/ Pron, Verb$_{to-be}$, Adj/ PastPart>:

(26) Press the key$_i$ down, and turn the volume up. Press it$_i$ again.

In MARS, this indicator was modified to use the information on noun-verb collocation from the whole document instead of only from one paragraph.

Immediate reference. This indicator is a specification of the collocation preference, and is highly genre specific, as the system is trained and tested on technical manuals that contain more imperatives than other genres. In constructions such as

... (You) V_1 NP ... con (you) V_2 it (con (you) V_3 it)

where $con \in \{and/or/before/after...\}$, the noun phrase immediately after V_1 is a very likely candidate for being the antecedent of the pronoun *it* immediately following V_2 and is therefore awarded +2 points.

(27) To print the paper you can stand the printer$_i$ or lay it$_i$ flat.

(28) To turn on the printer, press the Power button$_i$ and hold it$_i$ down for a moment.

(29) Unwrap the paper$_i$, form it$_i$ and align it$_i$, then load it$_i$ into the drawer.

Sequential instructions. [4] A score of +2 is assigned to NPs in the NP_1 position, which appear in the constructions of the form

To V_1 NP_1, V_2 NP_2 (Sentence). To V_3 it, V4 NP_4.

(30) To turn on the video recorder$_i$, press the button. To programme it$_i$, press the 'Programme' key.

Term preference. A score of +2 is assigned to NPs which represent the genre terminology, as they are more likely to be antecedents then NPs that are not terms. In MARS, which uses the whole text as anaphora resolution domain, the 10 NPs that appear with the greatest frequency in the document are considered significant, and all candidates matching one of the most frequent NPs are awarded the boosting score. This indicator is highly genre specific.

[4] This indicator was not mentioned in the original paper (Mitkov 1998), but was described later in Mitkov's book on anaphora resolution (2002).

Boost pronoun. In Mitkov's Original Approach, only nominal NPs were considered as candidates for antecedents. In MARS, previously resolved pronouns are also permitted to enter the list of antecedent candidates. The motivation for this addition is two-fold:

1. Pronominalized entities tend to be salient

2. The antecedent NP can be out of range of the algorithm and pronouns can thus be used as "stepping-stones" between the anaphor and the distant antecedent.

Since pronominal entries may reflect the salience of their antecedents, the pronominal antecedent candidates are awarded +1 point.

Syntactic parallelism. Contrary to Mitkov's Original Approach, MARS does have a syntactic parser, which makes it possible to take into consideration syntactic parallelism between an anaphor and its antecedent. The NP in the same syntactic role as the anaphor is awarded a score of +1.

Frequent candidates. This indicator is motivated by the observation that texts frequently contain a little group of entities being referred to most frequently by pronouns in the document. The boosting +1 is awarded to the three NPs that occur most frequently as antecedent candidates in the text.

Indefiniteness. The first of the impeding indicators, *Indefiniteness*, penalizes indefinite NPs with a score of -1, as they are less likely to be antecedents then the definite ones.

Prepositional NPs. The second of the impeding penalizes NPs appearing in prepositional phrases with a score of -1:

(31) Insert the cassette$_i$ into the VCR making sure it$_i$ is suitable for the length of recording.

In example 31, *the VCR* is penalized for being a part of *into the VCR*. This preference is based on the Centering theory's notion of salience. Brennan et al. (1987) propose ranking by obliqueness of grammatical relation of the subcategorized functions of the main verb: first the subject, then direct and indirect object followed by other subcategorized functions, and finally, adjuncts.

Referential distance. This indicator can impede or boost the chances that a candidate will be proclaimed as the antecedent. The indicator's effect is depending on the candidate's distance from the anaphor in terms of clause and sentence boundaries. In complex sentences, noun phrases in the previous clause (but in the same sentence) are the best antecedent candidates for an anaphor in the subsequent clause, and are awarded +2 points. Candidate NPs in the second previous sentence are awarded 1 point, in the sentence prior to that 0 points. In MARS, whose scope exceeds three sentences, all NPs in sentences further back are penalized with -1 points. For anaphora in simple sentences, noun phrases in the previous sentence are the best candidates for antecedents and are awarded 1 point. Noun phrases situated two sentences further back get 0 points and candidates three sentences back get -1 point. Different treatments of anaphora in complex and simple sentences is theoretical, as neither the implementation of MOA or of MARS have incorporated a clause splitter. In practice, only the latter approach is used.

It is important to keep in mind that antecedent indicators are *preferences* and not *constraints*, meaning that they can come into conflict by opting for different candidates. In example (31), the **prepositional phrase** indicator penalizes the *VCR* candidate giving it -1 point because it is a part of a prepositional phrase, but it does not destroy its chances of being chosen as antecedent as this negative score is overturned by the **collocation preference**, which awards it +2 points.

MOA achieved a success rate of 89.70%, while the fully automated MARS achieved 61.55%. I will discuss these results together with the results of my system in chapter 6.

3.2.2 The RAP system

Leass and Lappin's (1994) Resolution of Anaphora Procedure (RAP) is an algorithm for identifying inter- and intrasentential antecedents of anaphora in texts. It is applied to the syntactic representations generated by McCord's Slot Grammar parser (1990), and is, like the parser, implemented in Prolog. The discourse representation used by the algorithm consists of clausal representations of the previous four sentences in the text. The authors found the number of four sentences work best with technical texts. In addition it contains Prolog clauses declaring discourse referents evoked by NPs and specifying anaphoric links among the referents.

Leass and Lappin report 86% of succesfully resolved anaphora.

RAP contains seven main components that are first going to be shortly introduced, and then we will look at them in more detail:

1. An intrasentential *syntactic filter* that produces a list of pronoun-NP pairs in which referential dependence of the pronoun on the possible antecedent is excluded by syntactic constraints.

2. A *morphological filter* that eliminates antecedent candidates that do not match the anaphor in person, number or gender features.

3. A procedure for identifying pleonastic pronouns.

4. An anaphor-binding algorithm that identifies the possible antecedents of reciprocal or reflexive pronouns within the same sentence.

5. A procedure for assigning values to several *salience parameters* for an NP. The salience parameters of RAP correspond to the antecedent indicators in Mitkov's systems.

6. A procedure for identifying anaphorically linked NPs as an equivalence class, for which a global salience value is computed as a sum of the salience values of its elements.

7. A decision procedure for selecting the preferred antecedent candidate.

The syntactic filter

This filter is presented in a paper by Lappin and McCord (1990), and only a short overview will be given here. The filter covers the same domain as conditions B and C in Chomsky's binding theory (1981). It consists of six conditions for non co-reference between a pronoun and a non-reciprocal and non-reflexive NP.

A pronoun P is not coreferential with a noun phrase N if any of the following conditions holds:

1. P and N have incompatible agreement features.
   The agreement features of an NP are its number, person and gender features.

   (32) The woman$_i$ said he$_j$ was funny.

2. P is in the argument domain of N.
   Phrase P is in the argument domain of a phrase N iff P and N are both arguments of the same head.

   (33) She$_i$ likes her$_j$.

   (34) John$_i$ seems to want to see him$_j$.

3. P is in the adjunct domain of N.
   P is in the adjunct domain of N iff N is an argument of a head H, P is the object of a preposition PREP, and PREP is an adjunct of H.

(35) She$_i$ sat near her$_j$.

4. P is an argument of a head H, N is not a pronoun, and N is contained in H.

   (36) He$_i$ believes that the man$_j$ is amusing.

5. P is in the NP domain of N.
   P is in the NP domain of N iff N is the determiner of a noun Q and

   (a) P is an argument of Q
   or

   (b) P is the object of a preposition PREP and PREP is an adjunct of Q.

   (37) John's$_i$ portrait of him$_j$ is interesting.

6. P is the determiner of a noun Q, and N is contained in Q.

   (38) His$_i$ portrait of John$_j$ is interesting.

   (39) His$_i$ description of the portrait of John$_j$ is interesting.

The procedure for identifying pleonastic pronouns

This procedure is partially syntactic and partially lexical. The lexical part of the test contains a set of modal adjectives such as *necessary, possible, certain, likely, important, good, useful* etc., and a class of cognitive verbs: *recommend, think, believe, know, anticipate, assume, expect*. The pronoun *it* appearing in the following constructions is considered pleonastic:

- It is **Modaladj** that **S**
- It is **Modaladj** (for **NP**) to **VP**
- It is **Cogv-ed**[5] that **S**
- It seems/ appears/ means/ follows (that) **S**
- **NP** makes/ finds it **Modaladj** (for **NP**) to **VP**
- It is time to **VP**
- It is thanks to **NP** that **S**

[5] *Cogv-ed* is passive participle of cognitive verb

The anaphor binding algorithm

The following hierarchy of argument slots is defined:

subj > *agent* > *obj* > (*iobj*|*pobj*)

Subj is a surface subject slot, *agent* is the deep subject of a verb heading a passive VP, *obj* is the direct object slot, *iobj* is the indirect object and *pobj* is the object of the PP argument of a verb.

A noun phrase N is a possible antecedent for a reflexive or reciprocal pronoun P iff N and P do not have incompatible agreement features and one of the following conditions hold:

1. P is in the argument domain of N and N fills a higher argument slot than P.

   (40) They$_i$ wanted to see themselves$_i$.

   (41) Mary knows the people$_i$ who John introduced to each other$_i$.

2. P is in the adjunct domain of N.

   (42) He$_i$ worked by himself$_i$.

   (43) Which friends$_i$ plan to travel with each other$_i$?

3. P is in the NP domain of N.

   (44) John likes Bill$_i$'s portrait of himself$_i$.

4. N is an argument of a verb V, there is a noun phrase Q in the argument domain or the adjunct domain of N such that Q has no noun determiner, and

   (a) P is an argument of Q, or

   (b) P is an argument of a preposition PREP and PREP is in an adjunct of Q.

   (45) They$_i$ told stories about themselves$_i$.

5. P is a determiner of a noun Q and

   (a) Q is in the argument domain of N and N fills a higher argument slot than Q, or

   (b) Q is in the adjunct domain of N.

   (46) [John and Mary]$_i$ like each other$_i$'s portraits.

Salience weighting

Salience weighting is accomplished using *salience factors*. Each salience factor is associated to discourse referents in its scope, and is meant to reflect the salience of discourse elements associated with it. As the salience decreases with distance from the anaphor, for each new sentence that has been processed, the weight of salience factors is degraded by the factor of two. When the weight of a factor reaches zero, it is removed. The weight values are arbitrary and their function is to show comparative relations among the factors. Salience factors with their weights are given in table 3.1.

Factor type	Initial weight
Sentence recency	100
Subject emphasis	80
Head noun emphasis	80
Existential emphasis	70
Accusative emphasis	50
Non-adverbial emphasis	50
Indirect object and oblique component emphasis	40

Table 3.1: Salience factors in RAP

- The **sentence recency** salience factor is created for the current sentence and its scope is all discourse referents introduced in the current sentence.
- **Subject emphasis**
- **Head noun emphasis** rewards any NP not contained in another NP. Examples 47 and 48 are examples of NPs that do *not* get increased its salience value by this factor:

  (47) the assembly in *bay C*

  (48) the configuration information copied by *Backup configuration*

- **Existential emphasis** rewards predicate nominals in existential constructions:

  (49) There are only *a few restrictions* on LQL query construction for Wordsmith.

- **Accusative emphasis** rewards NPs that are direct objects.

- **Non-adverbial emphasis** rewards any NP not contained in an adverbial prepositional phrase. Examples 50 and 51 are examples of NPs *not* rewarded by this factor:

  (50) throughout *the first section of this guide*

  (51) in *the Panel definition panel*

- **Indirect object and oblique component emphasis**

Equivalence classes

Anaphor and *antecedent* are defined in terms of their relation that corresponds to DRT's (Kamp 1981) 'equality condition':

u = y.

where *u* is the discourse referent invoked by an anaphoric NP, *y* is a previously evoked discourse referent and = represents the anaphorical link. To avoid ambiguity with mathematical '=', Leass and Lappin use the word 'antecedes', so that the definition gets the following form

y antecedes u

Two discourse referents u and y are said to be co-referential, written as

coref (u,y)

if any of the following holds:

- y antecedes u
- u antecedes y
- z antecedes u for some discourse referent z and coref(z,y)
- z antecedes y for some z and coref(z,u)

In addition to this,

coref (u,u)

is true for any discourse referent u.

The equivalence classes of discourse referents are defined in terms of *coref* relation:

equiv (u) = {y| coref (u,y)}

All the members of an equivalence class belong to the same *anaphoric chain*. Each equivalence class has a a salience weight associated with it. This weight is the sum of the weights of all the salience factors that contain at least one member of the equivalence class in its scope.

The resolution procedure

RAP's procedure for identifying antecedents of pronouns is as follows:

1. Create a list of IDs for all NPs in the current sentence and classify them according to their type (definite NPs, pleonastic pronouns, other pronouns, indefinite NPs).

2. All NPs in the current sentence are examined.

   (a) NPs that evoke new discourse referents are distinguished from those that are presumably coreferential with already listed discourse referents.

   (b) Salience factors are applied to the discourse referents evoked in the previous step.

   (c) The syntactic filter and the first phase of the reflexive binding algorithm are applied.

      i. If the current sentence contains any personal or possessive pronouns, a list of pairs of IDs is generated. The list contains the pronoun-NP pairs for which coreference is ruled out on syntactic grounds. The syntactic filter that performs this step is described in section 3.2.2, p. 22.

      ii. If the current sentence contains any reciprocal or reflexive pronouns, each anaphor is paired with all of its possible antecedent binders. This is done by the anaphor binding algorithm described in section 3.2.2, p. 24.

   (d) Resolution of every non-pleonastic pronoun is attempted. The pronouns are resolved in order of occurrence in the sentence.

In the case of *reflexive* or *reciprocal pronouns*, the possible antecedents are identified by the anaphor binding algorithm. The candidate with the highest salience weight is chosen as the antecedent.

In the case of *third person pronouns*:

1. A list of possible antecedent candidates is created. Salience weights of the candidates are additionally modified, for example cataphora is penalized, while parallelism of grammatical roles is rewarded. These modifications are local to the resolution of a particular pronoun.

2. A salience threshold is applied: only the candidates with weights over the threshold are considered.

3. The possible agreement features such as number and gender for pronoun are determined.[6]

4. The morphological filter is applied

5. The syntactic filter is applied

6. If more then one candidate remains, the candidate with the highest salience weight is chosen. If several candidates have the same weight, the candidate closest to the anaphor is chosen.

7. The selected candidate is declared to be the antecedent of the pronoun.

3.3 Summary

In this chapter, I gave a short overview of the anaphora resolution field, followed by a closer presentation of two of the approaches: Mitkov's Original Approach (MOA) with MARS, and RAP by Lappin and Leass. Some of the rules applied by those systems have been implemented in ARN, and they will be discussed in greater detail in chapter 5.

[6]Parts of the algorithm cover languages with number ambiguity, such as in the Spanish possessive pronoun *su* or the German *sie*; Those parts will not be presented here.

Chapter 4

ARN: The architecture of the system

In this chapter I am going to describe the implementation of my system: Automatic Anaphora Resolution System for Norwegian – ARN. The chapter is divided into three main parts. In the first part I am going to give an overview of the data used, in the second I will describe how the data is organized, and in the third I will give an account of ARN's anaphora resolution module.

4.1 Data

The data used in this project includes two differently annotated versions of one corpus of Norwegian texts. Since both files are based on The Oslo Corpus of Tagged Norwegian Texts (Bokmål) I am going to call them 'the Oslo version' and 'the BREDT version'[1] in order to distinguish them. The main difference between the two sets of files is that the BREDT-files are manually tagged for referential chains.

The Oslo version

An example of an Oslo Corpus file entry can be seen in example 52[2]:

[1] The latter data set is obtained from the BREDT project of the University of Bergen, hence the name (http://bredt.uib.no/)

[2] The following tags have been used in this and the rest of the examples in this section:

- Syntactic: @subj *subject*, @obj *object*, @i-obj *indirect object*, @iv *infinite verb* @fv *finite verb*
- Morphological: verb *verb*, pres *present*, subst *noun*, mask *masculine*, ub *non-determined*, fl *plural*

(52)
```
"<venter>"  WD b1357=M b1358=M b526=S0 b730=S0 b771=S0
            "vente" verb pres @fv
            "vent" subst appell mask ub fl @obj @subj
```

Each entry starts with a single line that consists of the word token in square brackets (<>). Sometimes, the token is followed by the set of rules used by the tagger (`WD b1357=M b1358=M` etc. in ex. 52). If that is the case, the rules are simply ignored. If the word in question in unambiguous, the word token is followed by one line containing its lemma and its morphological, syntactic, and sometimes also semantic tags. If the word is ambiguous, as in example 52, several lines follow, each with its own lemma and its own sets of tags. The proper names consisting of several words are put together by the tagger and to some extent semantically tagged as in the example 53[3].

(53)
```
"<Jeremy A. Winther>"  RX b239=NM
           "Jeremy A. Winther" subst prop @subj &person &org
```

The BREDT version

In contrast to the entries in the Oslo version, the entries from the BREDT corpus always consist of only one line, containing the word's

1. serial/id number
2. token
3. ending
4. lemma
5. morphological class
6. syntactic function
7. numbers denoting referential chains, if any.

Some BREDT files entries are given in the following example:

(54)
```
b('23','gretne','tne','gretten','a','adj').
b('24','forretningsmenn','enn','forretningsmann','ni','pfill').
b('25','som','som','som','sbu','srel').
b('26','reiser','ser','reise','v','fv').
b('27','seg','seg','seg','pron','obj','23-24').
```

[3]The semantic tags used by The Oslo-Bergen tagger are: &person *person*, &org *organization*, &sted *location*, &verk *publication*, &hend *event* and &annet *miscellaneous*.

In this example, the word *seg* with serial number 27 is referring to the words *gretne* 'grumpy and *forretningsmenn* 'business men' with id-numbers 23 and 24, while the other entries in the example do not refer to the other words in the text.

Some minor formating changes have been conducted on the BREDT files in order to coordinate them with the corresponding files from the Oslo Corpus. The only major alteration is that the proper names consisting of more than one word were merged together, so that the entries such as the one in the example 55 are changed to entries such as in the example 56.

(55)
```
b( '78', 'Jeremy', 'emy', 'Jeremy', 'prp', 'n' ).
b( '79', 'A.', 'xprp', 'A.', 'prp', 'n' ).
b( '80', 'Winther', 'her', 'Winther', 'prp', 'subj' ).
```

(56)
```
b( '78', 'Jeremy A. Winther', 'her', 'Winther', 'prp', 'subj' ).
```

All the referential chain numbers in the files were changed according to the entry's new serial number.

Test corpus and training corpus

The Oslo-files and BREDT-files constitute a corpus of 15 pairs of equivalent files. File number 1, which covers approximately half of the corpus, is a collection of newspaper articles, while the rest of the files are excerpts from Norwegian literature. Each set of files contains 46972 words. The data is further divided into two parts of roughly the same size – a training corpus consisting of 21800 words and a test corpus of 25172 words.[4] The training corpus consists of files two through nine and the first 9916 words of file one, while the test corpus includes files ten to seventeen and the rest of file one. An overview of the size and genre of the files is given in table 4.1.

4.1.1 Limitations of data: Ambiguity

The main problem encountered in working with the Oslo Corpus files are unresolved ambiguities. There are especially three types of ambiguity that are central for anaphora resolution:

- Lexical ambiguity
- Semantic ambiguity in proper names
- Syntactic ambiguity

[4]The reason for this uneven division is that two of the files (files 3 and 9) that initially ended up in the training corpus turned out not to be manually tagged for referential chains. This was first discovered after the the corpus was divided.

File number	Number of words	Genre
1	19206	Newspapers articles
2	1969	Fiction
4	1988	
5	1995	
6	1984	
7	1970	
8	1978	
10	1988	
11	1993	
12	1976	
13	1998	
14	1976	
15	1978	
16	1966	
17	1988	
Total	46972	

Table 4.1: Data Files

Lexical ambiguity

As I have previously mentioned, ambiguous words, such as the one in example 52, have two or more lemmas followed by corresponding sets of tags. As solving this type of ambiguity would have exceeded the scope of this thesis, it has been left unsolved: In practice this means that if one of the meanings of an entry is noun, the entry has been treated as a noun.

Semantic ambiguity

The Oslo-Bergen tagger performs a semantic disambiguation of proper names. In cases where the tagger cannot determine the correct tag, several tags are left, such as in example 67, where the tags `&person` and `&org` imply that `Jeremy A. Winter` is either a person or an organization. This type of ambiguity has also been disregarded, i.e. if a noun has a `&person` tag, it is considered a person name.

Ignoring the ambiguities has, of course, its drawbacks, the main being that it impairs the results of the system. However, there are several arguments for this approach:

- Manual disambiguation of the corpus would be too time-consuming and would take time away from more central aims of this project.
- The system is now functioning with minimal human intervention.
- It would be easy to incorporate an improved corpus into the existing system.

Syntactic ambiguity

The unresolved syntactic ambiguity in example 57 has left the word *Pierre Chardin* with both `@obj`, `@subj` and `@i-obj` tags.

(57) Det dufter svakt Pierre Chardin av ham.
it smells weakly pierre chardin of him
He smells vaguely of Pierre Chardin.

Unresolved syntactic ambiguity such as this one would have posed a serious problem for ARN. However, the files in BREDT version have also been manually tagged with syntactic tags, so in case of semantic ambiguity, the data from BREDT files is used for disambiguation.

4.1.2 Other data sources

Apart from the two parallel corpora, five lists of nouns[5] have been used to determine whether a given noun denotes a human being:

1. A list of agents of temporary activity (133 words)
2. A list of agents of persistent activity (83 words)
3. A list of nouns covering kinship relations (66 words)
4. A list of professions (463 words)
5. A list of geographic origin and nationalities (1254 words)

Totally, the lists contain 1999 words. All antecedent candidates are checked against these lists, and whenever a match is found, it is marked within the respective word-object. The lists 1 - 4 contain three sublists each: one for words denoting persons of indeterminable sex, and two more for nouns denoting each sex. List number five consists only of words denoting persons of indeterminable sex. How the information from these lists has been incorporated into ARN will be described in the next section.

[5]The first four lists were obtained from the SIMPLE-project (http://cst.dk/simple/index.html), edited by Andra Björk Jónsdóttir (2003) and Lilja Øvrelid (2003). The fifth one is from the The Norwegian Language Council (Språkrådet) (www.sprakradet.no)

4.2 Organisation of data

The data is organized in two main structures:

- **Word-objects**, containing information on each word token
- **Sentence-objects**, consisting of lists of word objects

4.2.1 Word-objects

The process for making word-object is as following: The Oslo corpus file and the corresponding BREDT file are read simultaneously, and the information from them is merged to make word-objects. Simultaneously, ARN keeps track of data from the word lists (p. 33), and finally, in case of anaphora resolution ARN will complete the word object with data obtained from resolution. However, if the word is not of direct interest for anaphora resolution, such as e.g. the preposition *med* 'with', practically all the data needed to make the object can be found in the Oslo corpus word entry. We will first take a look at one such "simpler" type of word object and look into its main slots:

(58) The Oslo version entry:

```
"<med>"  WD b1414=M
        "med" prep @adv
```

(59) The word object:

```
Token: <med>
ID-number: -1
Reference ID: NIL
Lemma: (med)
Features: ((med) (prep @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
```

Explanation of slots

Token. This slot contains the word token as it appears in the text.

ID-number. If the word in question is not a noun or a third person pronoun, it gets the default value -1. Nouns and pronouns get their id-numbers from BREDT files where these numbers mark the word's position in the file.

Reference-ID. If the word does not refer to another word in the text, it gets the default value NIL, as in the given example. Otherwise, the slot is filled with data from the BREDT files. In the sample from the BREDT file, given as example 54 on page 30, the number in question is the last part of the entry of the word *seg*. The reference-id together with id-numbers is used in evaluation phase, where reference numbers of pronouns (anaphora) are compared to id-numbers of nouns (antecedent candidates).

Lemma. An unambiguous entry in Oslo corpus, such as the word *med* 'with', has only one lemma filled into this slot. If the word is ambiguous, all the lemmas are written here, unless they are identical in which case they are merged together. For instance, word *klubben* is the definite form of the words *klubb* 'club' and *klubbe* 'bat', so the word object's lemma slot contains both words.

(60)
```
Token: <klubben>
ID-number: 116
Reference ID: NIL
Lemma: (klubb klubbe)
Features: ((klubb) (subst appell mask be ent @$<$p-utfyll)
           (klubbe) (subst appell mask be ent @$<$p-utfyll)
(...)
```

Features. The `Features` slot contains all the data about an entry from the Oslo Corpus, except from the word's token. In other words, it contains lemmas and morphological, syntactic and semantic tags (if any). In the case of ambiguity, as in the example 60, it contains both sets of tags.

Previous & Next. These slots contain pointers the neighboring word objects.

As we have seen in example 59, if the word is neither a possible antecedent nor a possible candidate, the remaining slots have default value NIL. Before I give account of the rest of the slots, we can take a closer look at the word *hun* 'she':

(61) The Oslo version entry:

```
"<Hun>"  WD b3459=! b1401=M b495=S0H1 b1016=S!
       "hun" pron fem ent pers hum 3 nom @subj
```

The BREDT version entry:

```
b( '10548', 'Hun', 'hun', 'hun', 'pron', 'subj', '10541' ).
```

Final word object:

```
Token: <Hun>
ID-number: 10548
Reference ID: (10541)
Lemma: (hun)
Features: ((hun) (pron fem ent pers hum 3 nom @subj))
Previous: NIL
Next: #<WORD-TOKEN>
Antecedent: #<WORD-TOKEN>
     Antecedent's word-token:  <ELIN>
     Antecedent's ref-number: NIL
Points: ((10548 11 0) (10548 10 0) (10548 9 50) (10548 7 0)
         (10548 6 0) (10548 5 0) (10548 2 100) (10548 1 100))
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 10548.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 10548.
        Awarded 50 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 10548.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 10548.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 10548.
        Awarded 0 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 10548.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 10548.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 10548.
Extra features: #(1 0 4 10541)
                Sentence possition:
                    1
                The word's lemma appearing in the title:
                    No
                The word is denoting:
                    human being of female sex
                The pronoun is bound to the
                    antecedent with id-number 10541
```

The `ID-number` and `Reference ID` of this word object are obtained from its BREDT entry, given in the beginning of the example.

Antecedent. After the anaphor *hun* 'she' has been resolved, its antecedent is filled in the `Antecedent` slot, together with the antecedent's word token `(Elin)`. If the antecedent itself is an anaphor, its `Reference-ID` is also marked here. This data can later be used for making coreference chains.

Points. This slot is activated during the anaphora resolution process and keeps track of by which factors the antecedent candidate is given points during the resolution of different anaphora. The word in example 61 is an anaphor but it is also a possible antecedent for other anaphora. so its `points` slot is activated. This slot is followed by an explanation for convenience while browsing the file. We will look at the factors and the anaphora resolution process in greater detail in the next chapter.

Extra features. This slot is a vector of four elements that are initially set to zero. The slot is followed by an explanation that is not a part of the object.

- The first place in the vector is reserved for the word's sentence position, and is filled with numbers starting from one. This slot is activated under the anaphora resolution process, so that only words in sentences that contain anaphora get their sentence position marked.

- The second place in the extra-features vector is reserved for nouns only. If the noun, or more precisely, its lemma, has occurred in the section heading, the value of the second element of the vector is set to one. Otherwise, it remains zero.

- The third place is activated for nouns and pronouns and marks whether a noun/pronoun denotes a human being. It takes values from zero to four:

  - 0 - When the noun or the pronoun does not denote humans.
  - 1 - When the noun or the pronoun denote humans and we have no information on his/her sex.
  - 2 - When the noun denotes an human agent of male gender which can be of both male and female sex; Number 2 indicates that there is a slightly higher possibility that the noun denotes a human of male sex.
  - 3 - When the noun or the pronoun denote persons of male sex.
  - 4 - When the noun or the pronoun denote persons of female sex.

  The values for the nouns are set in two ways. If a noun has been tagged as a proper name, and also contains the tag `&person` with no extra information on the person's sex, it is assigned to class 1 "a human of unknown sex". In case of a known sex, the nouns are assigned as class 3 (males) or class 4 (females). If the noun is not a proper name, its lemma is checked against the contents of the five noun lists (section 4.1.2, p. 33) and if a match is found, the noun will be assigned to class between one and four. If no match

is found, the default value (zero) remains. We will return to sex/gender classification of nouns when I describe Factor 1 of the ARN system, in section 5.3.1, p. 47.

- In case of a successful anaphora resolution, the fourth place of the vector is filled with the antecedent's ID-number. Otherwise, the default value (zero) remains.

4.2.2 Sentence-objects

Anaphora are resolved within sentence-object, which consists of the sentence that contains the anaphor and the two preceding sentences. Sentence-objects consist of word-objects as described in the previous section and consist of four slots, as shown in example 62:

(62)
```
s-tokens:
(#<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN>)
  s-previous-1:
(#<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
 #<WORD-TOKEN>)
s-previous-2:
 (#<WORD-TOKEN> #<WORD-TOKEN> #<WORD-TOKEN>
  #<WORD-TOKEN> #<WORD-TOKEN>)
  stacks:
((100 #<WORD-TOKEN>) (100 #<WORD-TOKEN>)
 (100 #<WORD-TOKEN>) (100 #<WORD-TOKEN>)
 (50 #<WORD-TOKEN>) (50 #<WORD-TOKEN>))
```

The same sentence given with the word-objects' word tokens:

(63)
```
s-tokens:  (Han har reist seg fra sin stol på første
klasse og forbereder seg på møtet med norsk vinter .)
s-previous-1:( Langt fremme i flyet står Jeremy A. Winther .)
s-previous-2:  (Det er allerede mørkt .)
stacks:  ((100 vinter) (100 møtet) (100 klasse)
(100 stol) (50 Jeremy A. Winther) (50 flyet) )
```

The slot `s-tokens` contains the list of word objects of the current sentence, while the slots `s-previous-1` and `s-previous-2` contain the lists of word objects of the first and the second preceding sentences, respectively. In this way, the three sentences constitute a scope in which anaphora resolution is performed. When the next sentence is read, a new sentence object is made and the new-read sentence is placed in the `s-tokens` slot. The other two sentences (previously in `s-tokens`, and `s-previous-1` slots) move down one place while the sentence from the `s-previous-2` slot falls out of the scope:

(64)
```
s-tokens:  (Det dufter svakt Pierre Chardin av ham , en duft
han gleder seg over hver gang han løfter armene for å gjøre
en bevegelse .)
s-previous-1:  (Han har reist seg fra sin stol på første
klasse og forbereder seg på møtet med norsk vinter .)
s-previous-2:  (Langt fremme i flyet står Jeremy A. Winther
.)
s-stacks:  ((100 bevegelse) (100 armene) (100 gang) (100 duft)
(100 Pierre Chardin) (50 vinter) (50 møtet) (50 klasse) (50
stol) (0 Jeremy A. Winther) (0 flyet) )
```

If the sentence contains one or more anaphora in the `s-tokens` slot, the anaphora resolution module triggers a procedure that makes a list of antecedent candidates. During the course of anaphora resolution, those candidates are awarded points and the list of candidates together with their point scores is put in the `s-stacks` slot. At the end of the resolution of an anaphor, the candidate stack is stored, while the point scores are cleared, preparing space for a possible new resolution, which takes place if the current sentence contains more than one anaphor. A more detailed account on the anaphora resolution module is given in the next section.

4.3 The anaphora resolution module

The anaphora resolution module works in parallel with reading of files and the making of the word- and sentence-objects, and takes as input "freshly-made" sentence objects, as described in the previous section. It consists of four major steps:

1. Checking the sentence object for the presence of anaphora
2. Making a candidate stack
3. Application of resolution factors

4. Choosing the most appropriate candidate and evaluating the choice

We will first look more closely into each step, and at the end of the section I will give an example of resolution.

- Step 1: Finding the anaphor

  The first step in anaphora resolution is checking if the sentence object contains an anaphor.[6] If one or more anaphora are found, the list of anaphora to be resolved is made, and the resolutioner proceeds to the next step. If the sentence does not contain any anaphor, ARN proceeds to processing the next sentence object.

- Step 2: Making candidate stacks

  When a sentence contains anaphora, a stack of candidate antecedents is made, consisting of all the nouns and pronouns in the current sentence object. When all the candidates are gathered in the stack, the ones that have higher id-number than the anaphor, i.e. the ones that come after the anaphor in the text, are removed. In this way, ARN avoids attempting to solve cataphoric reference.

- Step 3: Applying factors

  In the process of resolving an anaphor, the factors are applied one at the time. Each factor gives rewarding (positive) or penalizing (negative) points to each of the candidates in the stack. In cases where a factor has no opinion on a candidate, zero points are awarded. These points are recorded two places: in the `s-stacks` slot of the current sentence object (section 4.2.2) and in the `Points` slot of the word object (section 4.2.1, p. 36). The points saved in the word object are saved together with information on the factors that have awarded the points, how many points are awarded, and the id-numbers of the anaphora that have been resolved.

- Step 4: Proposing and evaluating the antecedent

  When all the factors have been applied, the candidate stack is rearranged after score. The candidate with the highest point score is proposed as the antecedent if its score is higher than the given threshold, which is set to zero points. If none of the candidates' scores exceed the threshold, the resolution module gives the message that no appropriate candidates were found in the three sentence scope. If two candidates have the same number of points, the one closer to the anaphor is chosen. If an antecedent has been

[6] 'Anaphor' as defined in ARN's scope, p. 5. In the rest of the thesis, I will use words 'anaphor' and 'pronoun' in this sense.

proposed, its id-number is checked against the anaphor's reference number, and if the numbers match, the candidate is proclaimed correctly chosen. The anaphora resolution module keeps tracks of all the correct, incorrect and "no appropriate candidates" results for later evaluation of the system.

When the resolution of an anaphor is finished, next anaphor on the list is processed, and ARN returns to the step two of the algorithm. When the list is empty, a new sentence is read, the new sentence-object is made, and the resolution process begins anew.

I will now give an example (66[7]) of how ARN performs anaphora resolution on one sentence. This example is meant to give just a general idea of how ARN works. We will take a deeper look at ARN's factors, antecedent candidates and the resolution process in the next chapter.

As the example is pretty long, I will first give the sentence object as it would have looked like in normal text and its English translation, with the addition of id-number tags on its nouns and pronouns and the anaphor to be resolved in bold type:

(65) Hun$_{10516}$ hadde masse$_{10518}$ venner$_{10519}$, litt overalt, på samme måte$_{10526}$ som hun$_{10528}$ selv følte at hun$_{10532}$ ikke tilhørte noe spesielt sted$_{10537}$. Vi møtte Elin$_{10541}$ først ved inngangen$_{10544}$ til tenårene$_{10546}$. **Hun**$_{10548}$ hadde mot$_{10550}$ til å stå for sine ukonvensjonelle meninger$_{10557}$ og ideer$_{10559}$.

*She had lots of friends$_{10519}$, all over the place, the same way$_{10526}$ she$_{10528}$ felt she$_{10532}$ didn't belong to any particular place$_{10537}$. We first met Elin$_{10541}$ at the beginning of her teens$_{10546}$. **She**$_{10548}$ had the courage$_{10550}$ to stand for her own unconventional thoughts$_{10557}$ and ideas$_{10559}$.*

(66)
```
S-tokens:  Hun hadde mot til å stå for sine ukonvensjonelle
meninger og ideer .
S-previous-1:  Vi møtte Elin først ved inngangen til tenårene.
S-previous-2:  Hun hadde masse venner , litt overalt , på samme
måte som hun selv følte at hun ikke tilhørte noe spesielt sted.

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
```

[7]The English translations given in quotes are not the part of ARN's output.

```
The anaphor HUN has ten resolution candidates.
The numbers in front are their point scores.

187.5 Elin (id-nr 10541)          'Elin'
     Points: 100 pts by factor 1, 50 pts by factor 2,
     0 pts by factor 3, 50 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 0 pts by factor 9, 50 pts by factor 10,
     0 pts by factor 11, multiplied by 0.75

150.0 hun (id-nr 10532)          'she'
     Points: 100 pts by factor 1, 0 pts by factor 2,
     75 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 50 pts by factor 9, 0 pts by factor 10,
     0 pts by factor 11, multiplied by 0.5

150.0 hun (id-nr 10528)          'she'
     Points: 100 pts by factor 1, 0 pts by factor 2,
     75 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 50 pts by factor 9, 0 pts by factor 10,
     0 pts by factor 11, multiplied by 0.5

150.0 Hun (id-nr 10516)          'She'
     Points: 100 pts by factor 1, 0 pts by factor 2,
     75 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 50 pts by factor 9, 0 pts by factor 10,
     0 pts by factor 11, multiplied by 0.5

-37.5 tenaarene (id-nr 10546)      'teens'
     Points: -100 pts by factor 1, 50 pts by factor 2,
     0 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 0 pts by factor 9, 0 pts by factor 10,
     0 pts by factor 11, multiplied by 0.75

-37.5 inngangen (id-nr 10544)       'entrance (beginning)'
     Points: -100 pts by factor 1, 50 pts by factor 2,
     0 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 0 pts by factor 9, 0 pts by factor 10,
     0 pts by factor 11, multiplied by 0.75

-37.5 venner (id-nr 10519)          'friends'
     Points: -100 pts by factor 1, 0 pts by factor 2,
     0 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     0 pts by factor 7, 50 pts by factor 9, 0 pts by factor 10,
     -25 pts by factor 11, multiplied by 0.5

-62.5 maate (id-nr 10526)          'way'
     Points: -100 pts by factor 1, 0 pts by factor 2,
     0 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
```

```
     0 pts by factor 7, 0 pts by factor 9, 0 pts by factor 10,
     -25 pts by factor 11, multiplied by 0.5

-87.5 sted (id-nr 10537)          'place'
     Points: -100 pts by factor 1, 0 pts by factor 2,
     0 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     -50 pts by factor 7, 0 pts by factor 9, 0 pts by factor 10,
     -25 pts by factor 11, multiplied by 0.5

-87.5 masse (id-nr 10518)          'a lot of, lit. mass(n.)'
     Points: -100 pts by factor 1, 0 pts by factor 2,
     0 pts by factor 3, 0 pts by factor 5, 0 pts by factor 6,
     -50 pts by factor 7, 0 pts by factor 9, 0 pts by factor 10,
     -25 pts by factor 11, multiplied by 0.5
_____________________________________________________________
The proposed referent of anaphor HUN is ELIN

(wt-id reference): 10541
(wt-ref anaphor): (10541)

(CORRECT!)
-------------------------------------------------------------
Resolution of the sentence object completed
*************************************************************
```

The resolution output starts with the sentence object that has been processed. It is followed by the list of anaphora that are to be resolved. In our case, it is only one anaphor: hun_{10546} 'she'. The candidate stock is made, but before displaying it, all the factors are applied to all of the candidates and the applied factors are listed first. Subsequently, the candidate stack is displayed, with all the points each of the candidates has received. The list is rearranged after the candidates' point scores.

Data on each candidate consists of score point, its lemma, its id-number and the list of all the points that each factor has awarded the candidate. Finally, the candidate with the highest score (in our case Elin_{10541} with 187.5 points) is declared antecedent. The anaphor's reference number is checked against the antecedents id-number, and the resolution is pronounced (as in this case) success or failure.

4.4 Summary

The data ARN was trained and tested on comes from two versions of the Oslo-corpus files, one of which is manually tagged for referential chains. The sets

contain approximately 47 000 words each, and have been divided into a training and a test corpus.

The data is organized into word-objects, that contain all data on each word entry from both files, and sentence-objects that consist of word-objects from the sentence that contains the anaphor to be resolved and the two preceding sentences. During the resolution, ARN looks for antecedent candidates in this three-sentence window, applies rules (factors) to each of them, and finally proposes an antecedent. ARN then checks antecedent's id-number against the anaphor's reference number and declares whether the resolution was success or failure.

Chapter 5

ARN: The factors

The main part of this chapter will be devoted to the description of ARN's set of anaphora resolution rules (*factors*). Before we move on to the individual factors I will describe candidate lists, including the assignment of points and the types of word entities that can be included in candidate lists. In the second part of the chapter, the factors will be introduced, and described one by one. In that part we will also see why some of the factors that worked well for English are not well suited for an AR system for Norwegian. The chapter closes with a discussion of which factors should be included into the system.

All the results in this chapter come from the experimenting on the training corpus. The reason for this is that I did not want to apply ARN on the test corpus before the final version was ready. The results from the test corpus will be presented in the next chapter which deals with the final evaluation.

5.1 Candidates

Candidate lists consist of all nouns and third person pronouns[1] that precede the anaphor in question in the current sentence and all nouns and the third person pronouns in the two preceding sentences. Each factor rewards or penalizes each candidate with positive or negative points. When all factors have been applied, the candidate with the highest score is elected. If its point score exceeds a threshold, which is in ARN set at zero points, the candidate is proclaimed antecedent, and if not, it is announced that no appropriate candidates were found.

[1] Excluding *det* 'it (neut.)'.

5.2 Assigning points

There has been a broad agreement in the literature that range of the points that a system awards is arbitrary (Kennedy and Boguraev 1996 inter alia) and that what matters is the comparative relation among factors (Lappin and Leass 1994). Different anaphora systems have used different score ranges, e.g. from -1 to +2 (MARS by Mitkov 2002), from -5 to +5 (Pronominal anaphora resolution module of Lucy by Rich and LuperFoy 1988), and from 0 to 100 (RAP by Lappin and Leass 1994). ARN uses points in the range of -100 to +100.

The point scores are often modified by multiplying them with some other factor which could for instance reflect the candidate's proximity to the anaphor (Lappin and Leass 1994) or express the factors confidence (Rich and LuperFoy 1988). In ARN, the final candidate scores are multiplied by a number indicating their proximity to the anaphor. After some experimenting, I obtained best results when multiplying scores of the candidates in the current sentence with 1.0, in the penultimate sentence with 0.75, and scores of the candidates in the ante-penultimate sentence with 0.50.

5.3 Factors

I will first present a short overview of the implemented factors, and then discuss each factor at length. The following factors have been implemented in ARN:

- **Factor 1: Number/Gender/Animacy factor.**
  Most anaphora resolution systems have morphological constraints and preferences of some kind, so parallels to this factor can be seen in, among others, the checking for gender and number agreement that is done in the preprocessing stage of Mitkov's Original Approach (1998) (henceforth MOA), RAP's morphological filter (Lappin and Leass 1994) and in local anaphor constraints (Carbonell and Brown 1988).

- **Factor 2: The sentence proximity factor**.
  Corresponds to *Sentence recency* salience weight in RAP (section 3.2.2, p. 25) and the *Referential distance* factor in MARS (section 3.2.1, p. 21).

- **Factor 3: Boost pronoun.**
  Used under the same name by Mitkov's MARS system, described in section 3.2.1, p. 20.

- **Factor 4: Subject preference**.
  This factor has also been known as *First noun phrases*, *Givenness* and

Obliqueness in MOA and the MARS system (section 3.2.1, p. 17). It has also been applied in the RAP system under the name *Subject emphasis*.

- **Factor 5: Direct object preference**.
  The direct object preference is represented in both RAP and MARS; in the former under the name *Accusative emphasis* (section 3.2.2, p. 25). The factor did not appear in MOA, but in MARS it is, together with factors corresponding to ARN's factors 4, 6 and 7, a part of the *Obliqueness* factor.

- **Factor 6: Indirect object preference**.
  This factor figures in RAP and is a part of the *Obliqueness* factor of MARS.

- **Factor 7: Adverbial phrase penalization**.
  This factor is used by RAP under the name *Non-adverbial emphasis*. As RAP does not operate with negative points, this factor gives 50 points to NPs that are *not* part of adverbial phrases.

- **Factor 8: Prepositional phrase penalization**.
  In MARS the impeding factor *Prepositional NPs* penalizes the NPs that are part of prepositional phrases. The factor is described in section 3.2.1, p. 20.

- **Factor 9: Syntactic parallelism.**
  This factor does not appear in either RAP or MOA, but the NPs with the same syntactical role as the anaphor have been given precedence by the MARS system (section 3.2.1, p. 20).

- **Factor 10: Section heading preference**.
  The factor is applied by MOA and MARS and it is described on page 18.

- **Factor 11: Indefiniteness penalization**.
  The factor is applied by MOA and MARS (section 3.2.1, p. 20).

5.3.1 Factor 1: Number/Gender/Animacy factor

The question of gender and number is a complex one. Each noun has a grammatical number that can either be singular or plural, and can have any of the three genders - masculine, feminine and neuter, or the masculine/feminine (common) gender. Information about the grammatical features on nouns comes from the tagger, which uses information from a lexicon. There are, however, more factors than purely grammatical ones that influence the connection between a pronominal anaphor and the nominal antecedent. In many cases gender corresponds to sex, but that is not always the case. Similar problems, albeit in a much smaller scale, arise for singular and plural nouns in cases when singular nouns denote groups of

entities (e.g. *family*, *government*) and plural nouns denote single object (e.g. *scissors*). I have discussed gender and number at greater length in section 2.3.1, page 8, and will here focus on how ARN approaches this problem.

In many anaphora resolution systems this factor is implemented as a filter, discarding the candidates that do not match in gender and number. None of the papers I have mentioned in connection with this factor discusses it much. A typical example is Kennedy and Boguraev (1996) who mention in one sentence "a set of morphological filters which eliminate from consideration any discourse referent which disagrees in person, number or gender, with the pronoun". However, their result analysis shows that 35% percent of all the mistakes made by their AR system is due to gender mismatch, and one of their system improvement propositions is including a lexical data-base which includes detailed gender information.

I have chosen this approach – a lexical base that makes it possible to determine if a given noun denotes a person of male or female sex. The advantage of this approach is that we can avoid discarding the right candidate because of the gender/sex mismatch. Unfortunately, the database is too small, so many correct candidates are discarded anyway. This is more than a purely practical problem, since a database cannot theoretically ever be complete – language is not a static system that would be possible to map. But a bigger database would in any case do a better job.

The database used by ARN consists of six word lists[2] that together contain 1999 words:

1. A list of agents of temporary activity (133 words), containing words such as *beskytter* 'protector', *beslutningstaker* 'decision-maker', *kjøper* 'buyer', *taler* 'speaker', *øyenvitne* 'eye-witness'.

2. A list of agents of persistent activity (83 words), containing words such as *abonnent* 'subscriber', *alkoholiker* 'alcoholic', *bodybuilder* 'body-builder', *debattant* 'debater'.

3. A list of nouns covering kinship relations (66 words), containing words such as *tvilling* 'twin', *søster* 'sister', *svigermor* 'mother-in-law', *stesønn* 'stepson', but also words such as *mann* 'man', *dame* 'woman, lady' and *jente* 'girl'.

4. A list of professions (463 words), containing words such as *arbeider* 'worker', *astronaut* 'astronaut', *bonde* 'peasant', *designer* 'designer', *fisker* 'fisherman', *kulturminister* 'Minister of Culture', *visepresident* 'vice president'.

[2]The first four lists were obtained from the SIMPLE-project (http://cst.dk/simple/index.html), edited by Andra Björk Jónsdóttir (2003) and Lilja Øvrelid (2003). The the lists 5a) and 5b) are from the The Norwegian Language Council (Språkrådet) (www.sprakradet.no).

5. The last two lists denote the geographic origin of people from

   (a) Norway (817 words)

   (b) The rest of the world (437 words)

List one to four contain three sublists each:

1. A sublist for words denoting persons of indeterminable sex (*vitne* 'witness', *smugler* 'smuggler', *advokat* 'lawyer', *unge* 'kid').

2. A sublist containing nouns denoting persons of male sex (*sæddonor* 'sperm-donor', *pave* 'Pope', *stefar* 'stepfather').

3. A sublist containing nouns denoting persons of female sex (*politikvinne* 'police-woman', *servitrise* 'waitress', *tante* 'aunt').

List 5 a) and 5 b) consist only of words denoting persons of indeterminable sex.

In addition to the data from the data base, some semantic information is also obtained from the corpus, as the Oslo-Bergen tagger also includes some semantic tags. In example 67 tags `&person` and `&org` indicate that *Jeremy A. Winther* is either a person or an organization, but not, e.g. a work of art or a geographic location.

(67)
```
``<Jeremy A. Winther>'' RX b239=NM
    ``Jeremy A. Winther'' subst prop @subj &person &org
```

Based on the above-mentioned lists and on the additional data from the corpora, the nouns have been categorized in five groups according to their gender and animacy[3].

- **Class 1: Nouns that do not denote humans.**
  This class contains all nouns that are not identified as human proper names in the corpus, and that do not belong to any of the lists. As the nouns in this group do not have natural gender, their grammatical gender is taken into account, so that the class has three subclasses, masculine, feminine and neutral (*m*, *f* and *n* in table 5.1, p. 51).

[3]Animacy is not an unproblematic term. In its main sense, it denotes the attribute of being alive, as opposed to inanimate (non-living) objects. In a narrower sense, it denotes humans as opposed to both inanimate objects and living beings that are not humans. We will use the term in the latter, narrower sense.

- **Class 2: Nouns that denote humans of unknown gender, but which most probably are male.**
  I give the nouns in this class slightly higher odds that they are denoting persons of male than of female sex. The argument is merely that it's a man's world, and it is more probable that the person that is talked about is of male sex. This class contains sublists of lists one through four that denote humans of indeterminable gender and all the words in lists 5a) and 5b). In addition to these, this class also contains all nouns identified as human names of unknown gender, such as the case with most of the foreign names.

- **Class 3: Nouns that with certainty denote males.** This class consists of proper names identified as male or elements of list one to four that denote persons that necessarily are male, such as *bror* 'brother', *skjørtejeger* 'womanizer' and *baryton* 'baritone'.

- **Class 4: Words that clearly denote females.**
  This class includes proper names identified as female and members of the sublists of list one to four that contain nouns that are necessarily female, such as *fristerinne* 'temptress', *amme* 'wet-nurse' and *talskvinne* 'spokeswoman'.

According to the same standards, the third person pronouns can be described as:

- The pronoun that refers to non-humans: *den* 'it m./f.'.

- The pronoun that may or may not refer to humans: *de* 'they'/ *dem* 'them'[4] .

- The pronoun that refers to humans of unknown gender: *De* 'you (court.)/ *Dem* 'you (court. acc.)'.

- The pronoun that refers to males: *han* 'he'/ *ham* 'him'.

- The pronoun that refers to females: *hun* 'she'/ *henne* 'her'.

The pronoun for polite addressing (*De*) is not normally used in modern Norwegian, but it has been included since the corpus contains some older texts, written at a time when its usage was widespread. The third person pronouns with their forms are shown in table 2.1, on page 7.

The third person pronouns and the point sums awarded by factor 1 to the candidates belonging to the different classes are given in table 5.1.

[4]The pronouns are given here in their nominative and accusative form.

Pronouns			*den*	*de*	*De*	*han*	*hun*
Nouns	Class 1 *not human*	m	100	0	0	-100	-100
		f	100			-100	-100
		n	0			-100	-100
	Class 2 *human of indeterminable gender (male?)*		-50	75	75	100	75
	Class 3 *human male*		-50	75	75	100	0
	Class 4 *human female*		-50	50	50	0	100

Table 5.1: Factor 1: Points awarding

The points in table 5.1 cover the whole range from -100 to +100. The points were initially set according to grammar-based rules of the thumb, e.g. that while resolving the pronoun *hun* 'she', words that denote female persons should be given the maximal sum, while words denoting, say, non-living entities of neutral gender should be penalized. These rules were later experimentally adjusted.

We will now go column by column through the table 5.1:

- Under the resolution of anaphor *den* 'it', Factor 1 awards 100 points to the non-living entities of male and female gender, and it is indifferent towards the last subclass of class 1 – non-living entities of neutral gender. Although it penalizes nouns in class 2 through 4 with -50 points, it is actually also indifferent to them: replacing -50 with 0 made no difference in the training corpus. I have however decided to keep the -50 score to retain reciprocity to points given to non-living entities during the resolution of by anaphora *han* and *hun*.

- Under the resolution of anaphor *de* 'they', Factor 1 shows indifference towards the candidates from class 1 by giving them 0 points, while the candidates from the other classes are awarded 50 or 75 points. I did try to raise number of points given to the candidates from class 1, but it always impaired the results.

- As I have already mentioned, the third person singular pronoun for polite addressing (*De*) is on its way out from the Norwegian language. It has been used in two of the files of the training corpus that contain texts of older date. This pronoun does not refer to inanimate entities, but refers to both male and

female humans. For that reason, I started with penalizing the candidates from class 1 with -100 points, while the rest of the scores overlapped with those used for resolution of the pronoun *de* 'they'. This, however, impaired the results as did all the score combinations except the one presented in the table, which is identical with the scores used for the pronoun *de* 'they'. The reason for this could be the high incidence of ambiguity between these two pronouns.

- Under the resolution of pronoun *han* 'he', Factor 1 penalizes all the subclasses of class 1 with -100 points. Considering that my resources for deciding on the animacy of the noun are limited, penalizing candidates that were not found to denote humans with -100 points seemed too harsh. However, all the combinations where Factor 1 penalized candidates that were non-living entities of male gender with less then -100 yielded worse results.

- The situation with penalizing of candidates from the class 1 under resolution of anaphor *hun* 'she' is similar to that of the anaphor *han* 'he'. Experimenting with the scores given to candidates of class 2 turned out to be interesting: when Factor 1 awarded 75 instead of 100 points the candidates from this class, the results got slightly better, showing that this is, after all, a man's world.

The number issue is not taken into consideration in the distribution of points shown in this table. It is solved separately, in the following way: if an antecedent candidate has a number different from the anaphor in question, it is penalized with -100 points.

5.3.2 Factor 2: The reference proximity factor

As described in the beginning of this chapter, before a candidate is proclaimed antecedent, its final result scores are multiplied with 0.75 if the candidate belongs to the penultimate sentence and 0.50 if it belongs to the ante-penultimate sentence. The **Reference proximity factor** can be seen as an extension of this rule: All the candidates from the sentence that contains the anaphor in question are awarded 100 points, candidates from the penultimate sentence 50 points and candidates from the ante-penultimate sentence get no extra points. This is in compliance with both MARS and RAP, which both give the maximal point scores, +2 and 100 points respectively, to candidates in the same sentence (clause) as the anaphor, while the candidates in the penultimate sentence get +1 point from MARS' *Referential distance* and 50 points from RAP's *Sentence recency* weight.

5.3.3 Factor 3: Boost pronoun

Mitkov's original approach included only nominal NPs in the candidate lists, and the **Boost pronoun factor** was first included in MARS. As I have already outlined in section 3.2.1, Mitkov provides good argumentation for including this factor:

1. Pronominalized entities tend to be salient

2. The antecedent NP can be out of range of the algorithm and pronouns can thus be used as "stepping-stones" between the anaphor and the distant antecedent.

For this reason this factor was implemented in ARN, and the candidates that are pronouns are awarded 75 points. However, there was another, more practical, reason that this factor was introduced to ARN at an early stage, namely that the anaphora in BREDT are annotated in such a way that they form referential chains, rather then refering directly to the noun. Consider the following example:

(68) Den mørke vinterfrakken han tar på seg er myk og kler
the dark the-winter-coat he takes on [*him*]*self is soft and cloths*
hans slanke legeme, på samme måte som dressen, skjorten, slipset,
his slender body in same way as the-suit the-skirt the-tie
hatten, mansjettene og hanskene gjør det. Jeremy A. Winther$_{id.nr.159}$
the-hat the-cuffings and the-gloves do it jeremy a. winther
er vant til å se godt ut. Det eneste han$_{id.nr.172}$ ikke liker er
is used to to look well out the only he not like is
plastposen med konjakk og sherry som han$_{id.nr.182}$ holder i hånden.
plastic-bag with cognac and sherry that he holds in the-hand
The dark winter coat he is putting on suits his slender body, in the same way the suit, skirt, tie, hat, cuffings and gloves do. Jeremy A. Winther$_{id.nr.159}$ is used to looking good. The only thing he$_{id.nr.172}$ doesn't like is the plastic bag with cognac and sherry he$_{id.nr.182}$ is holding in the hand.'

In the BREDT files, *han*$_{id.nr.172}$ was marked as the antecedent of *han*$_{id.nr.182}$ and *Jeremy A. Winther*$_{id.nr.159}$ as the antecedent of *han*$_{id.nr.172}$. However, when ARN (correctly) chose *Jeremy A. Winther*$_{id.nr.159}$ as the antecedent of *han*$_{id.nr.182}$, this was rendered wrong, as it was not explicitly marked in the file that *han*$_{id.nr.182}$ also refers to *Jeremy A. Winther*$_{id.nr.159}$.

Introducing pronouns as antecedent candidates led again to a gender conflict. As we have argued in section 5.3.1, gender and number as nominal features can have somewhat blurred borders, but this is certainly not the case with pronouns.

The candidate list is therefore filtered right before pronouncing the most suitable antecedent, and the pronominal candidates of conflicting gender and number are removed.

This led to a slight increase of the results, but the number of correctly resolved anaphora also dropped in some files of the training corpus. This was due to the annotation of the corpus that allowed plural pronouns to refer to singular pronouns, such as in example 69.

(69) Hun hadde forresten merket helt siden han kom inn på det lille
she had by-the-way marked whole since he came in on the little
kontoret hennes den aller første dagen at han ikke likte henne.
the-office hers the very first day that he not liked her
Hvorfor han$_{id.nr.1020}$ ikke likte henne, var hun ikke klar over til å
why he not liked her was she not clear over to to
begynne med. Hun$_{id.nr.1035}$ hadde gjort sitt beste for at alt skulle gå
start with she had done her best for that all would go
fint mellom dem$_{id.nr.1047}$.
fine between them

Ever since he came to her little office on the first day, she noticed that he didn't like her. In the beginning, she was not aware of why he$_{id.nr.1020}$ disliked her. She$_{id.nr.1035}$ had done her best to make sure that everything would run smoothly between them$_{id.nr.1047}$.

In this situation, we could either keep the filter as it is, discard it, or use just a part of it, i.e. modify it in such a way that it filters only gender conflicting pronouns. The last alternative, although it brings some result increase seems like a poor, ad hoc, solution: deeming *hun* 'she' as a correct referent of *dem* 'them' does not seem right. If, on the other hand, the whole filter was discarded, we would risk having `'The proposed candidate for the pronoun SHE is HE'` resolutions, which is a grave mistake. There has not been written much about the types of mistakes an anaphora resolution system makes. Some mistakes seem less serious then the others: sometimes, human readers do not agree on the right antecedent, and if the anaphora resolution system chooses a candidate that has been the topic of discussion, but that was eventually not the antecedent the annotator deemed correct, that mistake cannot be considered serious. If, however an anaphora resolution system proposes the referent *hun* 'she' for the anaphor *han* 'he', that is a much more serious mistake. For those reasons I have decided to keep the number/gender pronoun filter in ARN.

5.3.4 Focus factors: factors 4, 5, 6, 7 and 8

Factors four through eight are all based on the same principle and will therefore be discussed as a group. The group includes the following factors:

- **Factor 4: Subject preference**
- **Factor 5: Direct object preference**
- **Factor 6: Indirect object preference**
- **Factor 7: Adverbial phrase penalization**
- **Factor 8: Prepositional phrase penalization**

All of these factors have in different forms been applied in both RAP and MARS. In MOA, which does not have a parser, the first NP, which is assumed to be a subject, gets +1 point, and in MARS, which has access to syntactic information, the subject NP gets the maximal +2 points, direct object gets +1 point, indirect object gets no bonus and the NPs which the parser cannot identify the function of are penalized with -1 (section 3.2.1, p. 17). Furthermore, NPs that are part of a prepositional phrase are penalized with - 1 point. Based on the Centering theory (Grosz, Joshi, and Weinstein 1995), Mitkov presents the following hierarchy of grammatical functions:

SUBJECT > DIRECT OBJECT > INDIRECT OBJECT > (OTHER).

In RAP, the subject NP gets 80 out of 100 points, direct objects are awarded 50 points, indirect object and oblique components get 40, and finally, NPs that are not a part of an adverbial phrase are also awarded 40 points. Lappin and Leass (1994) present the following hierarchy of argument slots:

SUBJECT > AGENT > OBJ > (IOBJ|POBJ)

Although Lappin and Leass explicitly state that they are not based on the Centering theory, it is difficult to ignore the similarity with the Centering theory's hierarchy of forward centers (C_f), as presented by Grosz et al. (1995):

SUBJECT > OBJECT(S) > OTHER

These factors are thus based on sound theoretical foundations, and are central in the systems ARN is based on. However, applying them in ARN, brought some unexpected results.

We will now look at each of the factors separately and try to find reasons for their unexpected behavior.

Factor 4: The subject preference factor

This factor was applied in ARN by awarding candidates that are in subject position 50 points. This impaired the results in all the files of the training corpus apart from two files whose number of correctly resolved anaphora remained the same. Awarding any other point sum led to worse results then awarding 50 points, but the best results are achieved by excluding this factor from the system.

I believe that the reason for that the failing of the **Subject preference** factor is to be found in the differences in information structure between the Norwegian and English languages. For reasons outlined in section 2.1, ARN does not resolve pronoun *det* 'it (neut.)', nor does it try to identify expletive subjects. Not being able to distinguish between logical and formal subject would perhaps not pose such a significant problem if it was not for the fact that Norwegian uses the expletive pronoun much more than English does, especially in the subject position, leaving many subjects unsuitable as reference candidates. The Norwegian Reference Grammar (Faarlund et al. 1997, p. 691) says that the subject is normally not the carrier of new information. In that, Norwegian does not differ from English. However, Norwegian goes to greater lengths go to avoid that a subject conveys new information. According to Faarlund et al. (1997, p. 1092) example 70 is not the natural answer to the question *Who found the money?*.

(70) Nils fann pengane.
nils found the-money
Nils found the money

The natural answer would be (71):

(71) Det var Nils som fann pengane.
it was nils who found the-money
Nils found the money

One way of keeping new information away from the subject position is indeed by using cleft construction as in (71): an expletive *det* 'it' is put in the place of subject when no other sentence parts can be used as subject. The cleft constructions in Norwegian also have the function of highlighting the information in the cleft clause, but Faarlund et al. (1997) consider this function secondary to the function of preventing the new information from becoming a subject. This may be a reason for the higher occurrence of cleft forms in Norwegian then in English. Further support for this idea is presented in Gundel (2002), who compared a Norwegian text with its English translation and came to the conclusion that clefts are much more commonly used in Norwegian then in English, as only 28% of Norwegian clefts were translated as clefts in English. Although she worked with a single text and a single translator, the results correspond to a more extensive study

of clefts, pseudo-clefts and inverted pseudo-clefts for Swedish and English (Johansson 2001). The study was based on 500 tokens of Swedish clefts and it found that only 33% were translated to English. As there seem to be no differences in distribution restrictions between Norwegian and English clefts, Gundel proposes that the reason for the more frequent use of clefts in Norwegian is that it shows a more consistent mapping between information structure and syntactic structure in making a clear distinction between topic and focus and between presupposed and non-presupposed content.

Besides clefts, constructions such as *presentational* and *topicalized* sentences are used for the same purpose. In contrast to cleft constructions where the formal subject is followed by the verb *være* 'to be', most of the one-place verbs can be part of a presentational construction, such as in example 72:

(72) Det arbeidet en mann i skogen.
it worked a man in the-wood
A man worked in the wood.

The exception is a small group of verbs that "*denote mental or sensory process that takes place inside a human being, like tenke* 'think' *or fryse* 'to be cold'. *A presentational focus must denote a situation that can be perceived in some way*" (Lødrup 1999). Yet another construction in which the expletive *det* 'it' appears and which has not been covered by either Gundel's or Johansson's examination is the impersonal passive construction. In contrast to both clefts (ex. 71) and presentational constructions (ex. 72), also verbs that take objects can have expletive *det* 'it' as a subject (ex. 73). In addition to direct objects, they can also take indirect object (ex. 74) or adverbial phrase (ex. 75) (Faarlund et al. 1997):

(73) Det vart fanga mykje fisk.
it was caught much fish
A lot of fish was caught.

(74) Det vart overrekt vinnaren ein pokal.
it became presented the-winner a cup
The winner was presented a cup.

(75) Det vart lagt ein dokument framfor oss.
it became laid a document in-front-of us
A document was laid in front of us.

The formal subject of the impersonal passive has the same semantic properties as the formal subject of presentational sentences, and cannot be the antecedent of an anaphor. The direct object must be in the indefinite form, but other NPs can

be finite. This is also interesting in connection to the **Factor 11: Indefiniteness penalization**, and we will come back to this subject when discussing that factor.

Factor 5: The direct object preference factor

In the Norwegian constructions that are used to avoid that the subject conveys new information, the role of bringing in the new information often falls on the direct object, such as in examples (71) and (72). This is another reason for favoring direct object NPs, on the top of also being supported by the centering theory. This factor did indeed raise the results of the training corpus, but the impact was weaker than expected: since the **Subject preference factor** fails, we could expect this factor to take over its impact and perhaps add some more, but including the **Direct object preference factor** into the system improved the results by only 0.29 percentage points, or 4 extra correct resolutions and one additional wrong resolution.

The single case where this factor excluded the correct candidate is given in the example 76:

(76) Vi har hatt drøftelser$_{id.nr.9395}$ med Viking$_{id.nr.9397}$, og vårt
we have had negotiations with viking and our
forlangende er en del høyere enn det de$_{id.nr.9408}$ tilbyr, sier
craving is a deal higher then that they offer says
klubbformann Gunnar Wilhelmsen i Tromsø til NTB.
club-president gunnar wilhelmsen i tromsø to ntb.

We had negotiations with Viking but what we ask for is somewhat higher than what they are offering, says Gunnar Wilhelmsen, the president of the Tromso club, to NTB.

During the resolution of the anaphor *de*$_{id.nr.9408}$ 'they' the correct antecedent, *Viking*$_{id.nr.9397}$, was chosen when Factor 5 was *not* applied, with the following point score:

100 Viking$_{id.nr.9397}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1

75 drøftelser$_{id.nr.9395}$ 'negotiations'
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
-25 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1

When Factor 5 joined the system, the candidate *drøftelser*$_{id.nr.9395}$ 'negotiations', got 50 points as the direct object, and came to the first place with 125 points. I do not, however, see this mistake as the failing of Factor 5, but as the failing of **Factor 1: The Number/Gender/Animacy factor**. If we look at the at the `features`-slot of the *Viking*$_{id.nr.9397}$ word object (77), we can see that it is correctly tagged as an organization:

(77)
```
Features:  ((Viking)
(subst prop @<p-utfyll &org &annet <org><annet>))
```

Had Factor 1 realized that organization can also denote a group of people (organization's members), Viking would end up in *Class 2 (Humans of indeterminable gender (male?))* and would be given 75 points, so that it would be chosen as antecedent no matter if Factor 5 was applied or not. Instead, Viking was not recognized as a group of humans, and it consequently fell into *Class 1 (not human)* getting 0 points (see table 5.1 on page 51). So why not just add a new rule to Factor 1? The situation is not that simple: the organization that is the football club Viking can denote a group of persons that are the organization's members, but other organizations, e.g. "Volvo" can, apart from denoting a group of persons also denote a car, a factory building, and much more. This problem area is discussed in work on name entity recognition (see e.g. Jónsdóttir (2003) for Norwegian and Mikheev et al. (1999) for English). It will not be focused on here, as name entity recognition is an area of its own that is too wide to be covered by this thesis.

Regarding the four anaphora that were correctly resolved when this factor was introduced to the system, all four were chosen instead of antecedents which belonged to the same referential chains. In other words, those resolutions were not actually wrong in the first place.

In conclusion, all I can say is that this factor was inconclusive - excluding or including it did not have any significant impact on the training corpus.

Factor 6: The indirect object preference factor

I have given the difference in information structure between Norwegian and English as the main cause of the different behavior of subject and direct object NPs in ARN, compared to the English-based RAP and MARS. As I see it, the only impact this difference could have on indirect object NPs is through the impersonal passive constructions that appear in Norwegian but not in English. There is, however, no reason why this construction should impair the indirect object NPs' salience. The previously given example of impersonal passive construction (78)[5] can be extended in two ways, as given in example (79) and in example (80).

(78) Det ble overrekt vinnaren en pokal.
it became presented the-winner a cup
The winner was presented a cup.

(79) Det ble overrekt vinneren en pokal. Den var to kilo tung
it became presented the-winner a cup it was two kilos heavy
og laget av ren gull.
and made of pure gold
The winner was presented a cup. It weighted two kilograms and was made of pure gold.

(80) Det ble overrekt vinneren en pokal. Han sa han var veldig
it became presented the-winner a cup he said he was very
glad fordi han hadde jobbet hardt for denne seieren.
happy because he had worked hard for this victory
The winner was presented a cup. He said he was very happy because he worked hard for the victory.

Examples 79 and 80 show that the salience of the indirect object (the cup) is at least as high as the salience of the agent (the winner). In spite of this, including this factor in ARN brought a minimal impairment of the results: the number of correctly resolved anaphora fell from 512 to 511, or from 73.89% to 73.74%, and this happened after I decided not to include **Factor 8: Prepositional phrase penalization** into the system. With Factor 8 included, the presence or absence of the Factor 6 did not make any difference.

[5]The extended examples are translated from nynorsk to bokmål as nynorsk has slightly different personal pronoun system.

The only case where including this factor led to an anaphor resolution that was deemed incorrect by the system was an example where the chosen candidate was a part of the same referential chain as the correct one. Thus the reason this resolution was rendered wrong was not the factor itself, but the evaluation system that should be developed to be able to keep track of longer referential chains. Why did this factor have such a negligible effect? There are two possible explanations:

- Awarding indirect object NPs points is a good strategy in half of the cases and counterproductive in the rest of the cases so that they cancel out.

- This factor is over-shadowed by other factors that award or penalize the same candidates that would otherwise be preferred by this factor, so that its influence is not felt. Those factors could for instance be **Factor 1: The Number/Gender/Animacy factor**, **Factor 2: The reference proximity factor** or **Factor 9: Syntactic parallelism**. Nevertheless, experimenting with removal of Factor 6 together with each of these factors did not give any results of interest.

I will come back to this factor in the last section of this chapter.

Factor 7: Adverbial phrase penalization

This factor is implemented in RAP, but as RAP does not operate with negative points, this was done indirectly, by giving 50 points to all NPs that are not a part of an adverbial phrase. The motive for introducing this factor is reasonable: to discourage candidates such as *uke* 'week' from examples such as (81):

(81) Jeg har venta på denne pakka lenge nå, og forrige
i have waited on this the-parcel(m./f.) long now and last
uke kom den endelig.
week(m./f.) came it(m./f.) finally

I have been waiting a long time for this parcel and last week it finally came.

Without this factor the anaphor *den* 'it (m./f.) would be resolved to *uke* 'week', since *uke* is the closest of the two candidates that match in gender and number.

In ARN, candidates that are parts of adverbial phrases are penalized with -50 points. This does not change ARN'a results, while all other point scores (except zero) impair the results.

I will explain why I nevertheless included this factor in ARN at the end of this chapter.

Factor 8: Prepositional phrase penalization

This factor is implemented in MARS with the purpose of discouraging candidates such as *stolen* 'chair' from the following example:

(82) Det lå en bok på stolen. Den var gammel og rød.
it lied a(m./f.) book on the-chair(m./f.) it(m./f.) was old and red.
There was a book on the chair. It was old and red.

In this case *den* 'it (m./f.)' must refer to *bok* 'book' although *stolen* 'the chair' is also an inanimate entity of the common (feminine or masculine) gender and can as such also be referred to by the pronoun *den* 'it (m./f.)'.

In MARS, this factor penalizes candidate NPs that are part of prepositional phrases with -1 point. However, experiments with this factor in ARN showed that the best results were achieved by giving a weak preference (25 points) to this type of candidates. When the factor was included in ARN and applied to the training corpus, the results were improved by 0.43 percentage points. An example where the correct antecedent was chosen by applying this factor is given in the following example:

(83) Hvor er kvinnen$_{id.nr.1592}$? Fra denne blokken i drabantbyen
where is the-woman$_{id.nr.1592}$ from this high-rise in the-suburb
forsvant hun$_{id.nr.1599}$. Bare hunden hun$_{id.nr.1603}$ var ute for å lufte
disappeared she$_{id.nr.1599}$ only the-dog she$_{id.nr.1603}$ was out for to air
er kommet hjem... Espen leser nøye om denne
is come home espen reads thoroughly about this
kvinnen$_{id.nr.1619}$ som har forsvunnet fra mann og barn
the-woman$_{id.nr.1619}$ which has disappeared from husband and children
natt til tirsdag, etter at hun$_{id.nr.1632}$ kom hjem fra sitt arbeid på
night to thursday after that she$_{id.nr.1632}$ came home from her work on
en bensinstasjon...
an gas-station
Where is the woman? She disappeared from this high rise in the suburb. Only the dog she was walking has come home... Espen reads carefully about this women who disappeared from her husband and children on Wednesday night, after she came home from her work at the gas station...

In this example, without the use of Factor 8, the best antecedent candidates for anapfor *hun*$_{1632}$ were as follows:

225.0 hun$_{id.nr.1603}$
Points: 100 pts by Factor 1: *Number/Gender/Animacy*

50 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
50 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 0.75
200 kvinnen$_{id.nr.1619}$
Points: 100 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1

When Factor 8 was added to the system, the candidate *kvinnen*$_{id.nr.1619}$ 'the-woman', got 25 points from Factor 8 and ended up sharing the first place with the candidate *hun*$_{id.nr.1603}$ 'she'. Since *kvinnen*$_{id.nr.1619}$ was closest to the anaphor, it was chosen as the antecedent, which was rendered correct by the system.

However, both candidates belong to the same referential chain: *hun*$_{id.nr.1632}$ refers to *hun*$_{id.nr.1603}$, which refers to *hun*$_{id.nr.1599}$, which refers to *kvinnen*$_{id.nr.1592}$, which is the same woman that we are talking about.

In all the cases where the inclusion of Factor 8 led to correct resolution, the newly proposed candidate belonged to the same referential chain as the old one. The improvement of the result can thus be seen as accidental and is more due to the fact that ARN is not equipped well enough for tracing referential chains than to the impact of Factor 8.

5.3.5 Factor 9: Syntactic parallelism

The candidates fulfilling the same syntactic role as the anaphor are awarded 50 points, which is in accordance with the corresponding factor in MARS that awards candidates +1 (out of max +2) points.

The inclusion of this factor into ARN led to an improvement of the system's performance in most of the files of the training corpus: in one file the results

remained unchanged, in two files the results weakened, and in four files the results improved. The overall improvement was 1.15% percentage points. This does not seem much, but it is one of the higher contributions of a single factor to the system. This factor's contribution was much higher in an older version of ARN where Factor 8 was also present. Including Factor 9 into that system improved the result by 4.47 percentage points.

Most of the the new candidates that were proposed when this factor was introduced to ARN did not belong to the same referential chain as the previously proposed candidates. This difference in resolution can be seen as significant, as opposed to most of the cases of the new reference proposition that was brought to by including the "focus factors", where the new and the old candidates belonged to the same referential chain. Of 14 cases where resolution was successful due to this factor, ten were significant, while this was the case with only two of the six cases where including Factor 9 led to wrong resolutions. I will now give a few examples where introducing this factor led to correct results. In most of the cases, both the anaphor to be resolved and the candidate that was awarded by Factor 9 had the function of subject, while in the remaining cases both were prepositional phrase supplement. The example (84) houses both:

(84) Selv skulle de fortsette livet i sine barn, og de$_{id.nr.879}$
(them)selves would they continue the-life in their children and they
ville skape trygghet for dem$_{id.nr.884}$, en trygghet som de$_{id.nr.889}$ ikke
should create security for them a security which they not
selv hadde fått.
(them)selves had got

They would continue their lives in their children, and they$_{id.nr.879}$ would create security for them$_{id.nr.884}$, a security they$_{id.nr.889}$ never got themselves.

Let's first look at the resolution of the anaphor *dem*$_{id.nr.884}$ 'them' with features `((de) (pron fl pers 3 akk @<p-utfyll))`. Without implementing Factor 9, the chosen antecedent was *de*$_{id.nr.879}$ 'they', which in itself is referring to (what we suppose are) the parents, and not the children. The top of the candidate list is as follows:

175 de$_{id.nr.879}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*

0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1
175 $de_{id.nr.871}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1
150 $barn_{id.nr.876}$
Points: 75 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 10: *Section heading preference*
-25 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1

The correct antecedent, *barn* 'children', is on the third place, but as it has the same syntactical tag as the anaphor `(@<p-utfyll)`, it gets 50 points when the Factor 9 is implemented, thus moving to the top of the antecedent candidate list with 200 points.

When the anaphor $de_{id.nr.889}$ 'they' with features `((de) (pron fl pers 3 nom @subj))` is resolved without Factor 9, the anaphor is resolved to $dem_{id.nr.884}$. This candidate is itself an anaphor which, as we have seen, refers to *barn* 'children', while $de_{id.nr.889}$ refers to *foreldre* 'parents'. The top of the antecedent candidate list is as follows:

175 $dem_{id.nr.884}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*

0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1
175 $de_{id.nr.879}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1
175 $de_{id.nr.871}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1

With introducing Factor 9, the correct antecedent, $de_{id.nr.879}$ with features `((de) (pron fl pers 3 nom @subj))` gets 50 points and reaches the first place with 225 points.

5.3.6 Factor 10: Section heading preference

In the training corpus, this factor is only applicable to file number one, as it is the only part of the corpus that contains section headers (it consists of newspapers articles). Although it is by far the biggest file in the corpus, covering approximately half of it, it is also the file with fewest third person pronouns compared to the size, The third person pronouns constitute 1.16% of the total number of the words, compared to 4.87% in the rest of the files. The file thus contains a high percentage of sentences and even some whole newspaper articles without any third person pronouns. This could explain the extremely low impact of this factor: only

one extra anaphor was resolved when it was introduced. This single resolution, shown in example (85), is nevertheless exactly what I wanted from this factor. The sentence in the example comes from an article titled *"Gambler ikke med Shearer"* 'Not gambling with Shearer':

(85) Den andre spissen Les Ferdinand$_{id.nr.7948}$ brakk kinnbeinet i
the other forward les ferdinand broke the-cheekbone in
helgens serierunde, og dermed var det fristende å forsere
the-weekend's match and therefore was it tempting to speed-up
Shearers$_{id.nr.7963}$ comeback. – Jeg gambler ikke. Han$_{id.nr.7971}$ må få
shearer's comeback i gamble not he must get
noen treninger før han spiller, og jeg har mange spillere i
some trainings before he plays and i have many players in
stallen, sa Keegan mandag.
the-stable said keegan monday

The other forward Les Ferdinand broke his cheekbone in a match this weekend, and it was therefore tempting to speed up Shearers comeback. - I am not gambling. He$_{id.nr.7971}$ needs to have some trainings before he plays, and I have many good players in reserve, said Keegan on Monday.

Before Factor 10 was introduced, the system resolved *Han*$_{id.nr.7971}$ 'he' to *Les Ferdinand*, with the following situation on the top of the candidate stack:

75.0 Les Ferdinand$_{id.nr.7948}$
Points: 100 pts by Factor 1: *Number/Gender/Animacy*
0 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
50 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 0.5
50.0 Shearers$_{id.nr.8514}$
Points: 100 pts by Factor 1: *Number/Gender/Animacy*
0 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*

0 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 0.5

With the introduction of Factor 10, *Shearers* 'Shearer's' got 50 points thus sharing the top position with *Les Ferdinand*, and was chosen the correct antecedent being the closest of the two.

5.3.7 Factor 11: Indefiniteness penalization

RAP does not use this factor, but MARS punishes indefinite NPs with the maximal penalty of -2 points. In ARN this factor penalizes the indefinite NPs with -25 points, but it brings a minimal improvement of 0.3 percentage points. What are the reasons for such a small impact? We will first look at why this rule was introduced in the first place:

(86) Gutten sparka en ball. Han var sinna og kikket rundt etter flere
the-boy kicked a ball he was angry and looked around after more
ting han kunne sparke.
thing he could kick

The boy kicked a ball. He was angry and looked for more things he could kick.

In examples such as (86), the NP in definite form is the theme of the discourse, and penalizing the indefinite NP is the correct strategy. Consider the following examples, however:

(87) Det kom en mann gående nedover stien.
it came a man going down the-path

A man came down the path.

(88) Han var høy og mørk.
he was tall and dark.

He was tall and dark.

(89) Den var nesten helt gjengrodd.
it was almost entirely re-grown

It was almost entirely overgrown.

Example (87) is a presentational construction, and the expected continuation would be (88), not (89). So, although *en mann* is in the indefinite form, it is him the discourse is about. Similarly, the discourse is obviously not about the road, although *veien* 'the road' is in the definite form, and penalizing the indefinite NP or promoting the definite NP would favour the wrong candidate. I have earlier in this chapter (p. 56) mentioned the high appearance of expletive *det* 'it' subjects in Norwegian. They are used to avoid introducing new information by subject, and they appear much more often in Norwegian then in English.

I will come back to some examples of sentences including expletive *det* 'it' which I gave earlier:

(90) Det er ein flekk på kjolen din.
it is a stain on the-dress yours

There is a stain on your dress.

(91) Det arbeidet en mann i skogen.
it work a man in the-wood

A man works in the wood.

(92) Det vart overrekt vinnaren ein pokal.
it became presented the-winner a cup.

The winner was presented a cup.

In all the expletive *det* 'it' constructions I have written about earlier in this chapter, i.e. topicalization (ex. 90), presentation (ex. 91), and impersonal passive (ex. 92), the direct object has to be in the indefinite form. The other NPs such as indirect object, and NPs that are part of prepositional or adverbial phrases can be in definite form.

The number of correctly resolved anaphora in the training corpus increased by two after the introduction of Factor 11: four additional resolutions were done correctly, but in two the correct candidate was discarded. In the following example, the correct candidate was discarded:

(93) Det viser seg at disse store bedriftene bruker alle de
it shows (it)self that these big the-corporations use all the
pengene de$_{id.nr.8491}$ får fra Forskningsrådet, og egne midler i
money they get from the-research-council and own resources in
tillegg, til å kjøpe tjenester i forskningsmiljøene. Når det
addition to to buy services in the-research-communities when it
hevdes at flere forskningsmiljøer$_{id.nr.8514}$ mener at dette er gal
claims that more research-communities mean that that is wrong

bruk av knappe, offentlige midler, må de$_{id.nr.8528}$ være klar over
use of scarce public resources must they be aware over
at disse midlene er med på å utløse ytterligere midler fra
that these the-resources are with to to free further resources from
bedriftene selv til å gjennomføre forskning i
the-companies (them)selves to to carry-out research in
forskningsmiljøene.
the-research-communities

It turns out that the big companies use all the money they$_{id.nr.8491}$ get from the Research Council, and their own means in addition, to buy services from the research communities. When it is claimed that several research communities$_{id.nr.8514}$ think that this is the wrong usage of scarce public resources, they$_{id.nr.8528}$ have to keep in mind that those resources release further resources from the companies that are used for research in the research communities.

The anaphor *de*$_{id.nr.8528}$ in this rather long and complicated example was resolved correctly when Factor 11 was excluded. The correct candidate, *forskningsmiljøer*$_{id.nr.8514}$ 'research communities' had the same total score as the nearest rival candidate, *de*$_{id.nr.8491}$ 'they', but was chosen as it was closest. When the Factor 11 was applied, the -25 points it penalized the correct candidate with made it fall to the second place:

150.0 de$_{id.nr.8491}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
50 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
50 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
0 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 0.75
125 forskningsmiljøer$_{id.nr.8514}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*

50 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
-25 pts by Factor 11: *Indefiniteness penalization*
The point score is multiplied by 1

In this case, the two candidates do not belong to the same referential chain, as *de*$_{id.nr.8491}$ refers to *bedriftene* 'the companies' and not *forskermiljøer* 'research communities'.

The training corpus contains four more cases where the correct candidate was chosen thanks to this factor, but one of those four is insignificant, as the candidates belong to the same referential chain.

In the example 94 one of the three significant correct resolutions is given:

(94) For de hadde alle vært igjennom det samme og var ikke lenger
for they had all been through the same and were no longer
fremmede. De$_{id.nr.959}$ hadde fått et fellesskap av de smerter og
strangers they had got a brotherhood of the pains and
savn som var påtvunget dem$_{id.nr.972}$. Men mange broer skulle
longing that was forced them but many bridges would
de$_{id.nr.978}$ bygge til hverandre.
they build to each-other

As all of them had been through the same, they were no longer strangers. They$_{id.nr.959}$ had a brotherhood that rose from the pain and longing that was forced upon them$_{id.nr.972}$. But they$_{id.nr.978}$ were going to build many bridges towards each other.

During the resolution of anaphor *de*$_{id.nr.978}$ 'they', the proposed candidate before introducing Factor 11 was *broer* 'bridges':

150.0 broer$_{id.nr.976}$
Points: 100 pts by Factor 1: *Number/Gender/Animacy*
100 pts by Factor 2: *The sentence proximity*
0 pts by Factor 3: *Boost pronoun*
50 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
50 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
The point score is multiplied by 1
150.0 dem$_{id.nr.972}$

Points: 0 pts by Factor 1: *Number/Gender/Animacy*
50 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
50 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
0 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
The point score is multiplied by 0.75
150.0 $\text{De}_{id.nr.959}$
Points: 0 pts by Factor 1: *Number/Gender/Animacy*
50 pts by Factor 2: *The sentence proximity*
75 pts by Factor 3: *Boost pronoun*
0 pts by Factor 5: *Direct object preference*
0 pts by Factor 6: *Indirect object preference*
0 pts by Factor 7: *Adverbial phrase penalization*
50 pts by Factor 9: *Syntactic parallelism*
0 pts by Factor 10: *Section heading preference*
The point score is multiplied by 0.75

Factor 11 penalized the indefinite *broer* 'bridges' with -25 points, so it came to the last place, while the correct candidate $dem_{id.nr.972}$ 'them' came to the first place sharing the same number of points as the second-placed candidate $De_{id.nr.959}$ 'they' which belongs to the same referential chain.

This factor has both some advantages and some drawbacks and its overall contribution to the system is rather small. The situation with Factor 11 is similar to that of the "focus-factors": The factor's contribution is unclear when it is applied on this, rather small, set of data, so more data and more experiments are needed for a more thorough evaluation of the factor.

5.4 Which factors to include in ARN?

Before applying ARN to the test corpus, I had to decide what factors to include in the system. I will be using the following two criteria when evaluating each factor's importance:

- Criterion 1: How many correct resolutions does a factor accomplish when applied alone?
- Criterion 2: How much does a factor contribute to the system i.e. how much does ARN lose in performance when the factor in question is removed?

However, as none of the criteria is decisive, they will be supplemented with analysis of examples in which the factors contributed to the correct or incorrect resolutions. I will also take into consideration theoretical foundations for each factor.

Factor	Criterion 1 (%)	Criterion 2 (percentage points)
Factor 1: Number/Gender/Animacy factor	66.67	12.70
Factor 2: The sentence proximity factor	26.12	4.48
Factor 3: Boost pronoun	60.32	3.32
Factor 4: Subject preference	47.91	-2.74
Factor 5: Direct object preference	17.32	0.29
Factor 6: Indirect object preference	26.27	-0.14
Factor 7: Adverbial phrase penalization	27.71	0.00
Factor 8: Prepositional phrase penalization	12.84	0.43
Factor 9: Syntactic parallelism	44.88	1.16
Factor 10: Section heading preference	26.12	0.15
Factor 11: Indefiniteness penalization	41.41	0.29

Table 5.2: The factors according to Criteria 1 and 2

We will start with considering each factor, one by one.

Factor 1: Number/Gender/Animacy factor

Applied alone, this factor resolves correctly 66.67% anaphora and its contribution to the system is by far the highest (12.70). Its performance is excellent and its theoretical basis sound, so this factor is included in the system without much further discussion.

Factor 2: The sentence proximity factor

This factor has a moderate percentage of correct resolution when applied alone (26.12%), but its relatively high contribution to the system (4.48) makes it a strong candidate.

Factor 3: Boost pronoun

Factor 3, with 60.32% on Criterion 1 and 3.32 percentage points on Criterion 2, scores high on both criteria, and this is, together with its sound theoretical basis, sufficient reason to include it in the system.

Factor 4: Subject preference

Although Mitkov argues strongly for subject preference and Lappin and Leass give subjects very high salience weight, the situation concerning this factor is not at all straight forward. I have argued earlier in this chapter (pp. 56-58) that this factor has worse performance for Norwegian compared to English due to differences in the information structure. Furthermore, no matter which point score this factor rewards or penalizes candidates with, it always impairs the results – the least damage it does is lowering the ARN's performance by 2.74 percentage points. We can conclude that in spite of the high percentage of correctly resolved anaphora when the factor is applied alone (Criterion 1 = 47.91%), there are two strong reasons for excluding it:

- There is a good theoretical explanation for its failure.

- Its presence impairs the system's performance.

Those arguments weigh so much that I have decided to exclude the factor from the ARN.

There are, however, further complications in connection with the evaluation of this factor, which are highly relevant for deciding whether or not to keep the rest of the "focus-factors"[6] ARN. The problems in question are:

- **The problem of *det* 'it (neut.)'**
  ARN does not resolve pronoun *det* 'it (neut.)', and does not try to disambiguate it from the expletive subject of the same form, for reasons outlined in the introductory chapter (p. 5). Since we do not know how many of the subjects are expletive *det* 'it', we do not know anything about the distribution of neither presentational and topicalized constructions, nor impersonal passives. This distribution does not only influence the subject's suitability as potential antecedent, but also the performance of all the Centering theory-based factors, i.e. Factors 5, 6, 7 and 8 and, to some extent, Factor 11.

- **Reference chains**
  If the proposed candidate is itself an anaphor, ARN can check if the candidate has the same antecedent as the anaphor, and, in a case of match, proclaim the resolution correct. This means that if anaphor a_1 refers to anaphor a_2, and a_2 refers to noun n_1, then if the anaphor a_1 is resolved to the noun n_1, this resolution is rendered correct, although the referential bond a_1-n_1 is not tagged in the BREDT-files. In this way, ARN makes short referential chains

[6] Discussed earlier in this chapter, on pp. 55-63

of three links ($a_1 \rightarrow a_2 \rightarrow n_1$). The limited length of these chains poses a problem in the evaluation of the factors' impact on the systems, as there are not few examples where applying a factor led to apparently incorrect resolution, but the antecedent that was deemed incorrect and the antecedent that was manually tagged as correct in the BREDT-files belonged to the same referential chain.

Factor 5: Direct object preference

This factor is the first of a long line of factors that score low or moderately on Criterion 1, and that score exceptionally bad on Criterion 2. Factor 5 itself scores low according to both criteria: applied alone, it resolves 17.32%, and when it is removed from ARN, the system's results fall by only 0.29 percentage points. When it comes to the examples of correct resolutions when this factor was applied, there were not any significant ones (see sec. 5.3.4, p. 58). I have decided to keep this factor for purely theoretical reasons: If the subject NPs fail as antecedent candidates, there is a good reason to believe that object NPs will be playing a more prominent role as antecedents in Norwegian than they do in English. I would expect this factor to perform better once the problem of *det* 'it (neut.)' is solved and when longer referential chains are handled. In other words, I am keeping the factor for an improved version of ARN.

Factor 6: Indirect object preference

This factor is similar to Factor 5: it scores relatively low on Criterion 1, 26.27%, and it actually scores bellow zero (-0.14) on Criterion 2, as the number of correctly solved anaphora falls from 512 to 511. I have shown in the previous section (p. 60) that this anaphor should not be considered wrongly resolved as the chosen antecedent belongs to the same referential chain as the one tagged as correct. The slight decrease in ARN's performance is thus considered accidental and not due to this factor's interference. Having that in mind, the reasons for keeping this factor are the same reasons as for keeping Factor 5: I believe that it has potential to perform better when issues of expletive subject *det* 'it' and referential chains are solved.

Factor 7: Adverbial phrase penalization

The **Adverbial phrase penalization** factor builds on the same principle as the two previous factors, i.e. the Centering theory (p. 56). Its performance according to Criterion 1 is mediocré (27.71%) and, according to Criterion 2, bad: 0 percentage points, i.e. it apparently does not influence the results at all. I do not, however,

exclude the possibility that the number of correctly resolved anaphora was neutralized by the number of wrongly resolved. I have decided to keep this factor for the same reasons as for Factors 5 and 6.

Factor 8: Prepositional phrase penalization

Factor 8 performs poorly according to Criterion 1 (12.84%), but it scores relatively well on Criterion 2, increasing ARN's performance by 0.43 percentage points. Factor 8 has been excluded from ARN, although it scored better on Criterion 2 than both factors 5, 6 and 7.

In the beginning of this section I emphasized that Criteria 1 and 2 are not decisive, but can rather be seen as indicators of the factors' performance – some of the factors presented here have been included in ARN in spite of their poor performance on one or both criteria, as long as they had a solid theoretical background and did not harm the system.

The **Prepositional phrase penalisation factor** is based on the same principle as the rest of the "focus factors". The failing of Factor 4 has been explained in terms of information structure differences between Norwegian and English. This difference, expressed in higher occurrence of sentences with expletive *det* 'it' in Norwegian, also influences for the rest of the members of this factor group. It is, however, difficult to assume that these constructions rise the salience of prepositional phrases so strongly that these candidates, from being strongly discouraged in English, become mildly preferred antecedent candidates in Norwegian. Thus I do not find any explanation as for why giving week preference to prepositional phrases improves the result, other then that is accidental. Examples of correct resolutions that made the 0.43% increase in ARN's results confirm this theory: in all of the cases where including Factor 8 led to corrct resolution, previously chosen candidates belonged to the same referential chains. Not finding any theoretical explanation for Factor 8's apparent success was decisive in not including it in ARN.

Factor 9: Syntactic parallelism

Factor 9 scores fairly well on both criteria: 44.88% on Criterion 1 and 1.16 percentage points on Criterion 2. It has a good theoretical background, gives a solid contribution to the system and examples of successful resolutions after the factor was introduced were significant, i.e the old and the new-proposed antecedents did not come from the same referential chain. This factor is therefore included in ARN.

Factor 10: Section heading preference

The corpus on which this factor was trained was much smaller then the corpus for other factors. Although by its application only one correct resolution was achieved, the factor performed exactly what was expected of it. In addition, the proposition that entities contained in the titles of articles have increased salience is solid, so Factor 10 is kept.

Factor 11: Indefiniteness penalization

Factor 11 resolves 41.41% when applied alone (Criterion 1), but it contributes with only 0.29 percentage points to the system. The salience of indefinite noun phrases is closely related to the question of expletive *det* 'it', almost as closely as the question of salience of subject phrases, objects, adverbial and prepositional phrases (Factors 4, 5, 6, 7 and 8). In all of 'det'-constructions: presentations, topicalizations, clefts, and impersonal passives, the salience of indefinite NPs rises slightly, but it is unclear if there are enough examples of these constructions and if the rising of salience is strong enough to annulate the rest of the cases where indefinite NPs are less salient than definite. While waiting for a better data set, this weak penalization is kept in ARN.

Conclusion

The 11 considered factors can now be divided into three groups:

- The four factors that constitute the "hard core" of the system, whose inclusion went without question, based on their relatively good C2 scores and firm theoretical background:
  - Factor 1: Number/Gender/Animacy factor
  - Factor 2: The sentence proximity factor
  - Factor 3: Boost pronoun
  - Factor 9: Syntactic parallelism
- Two of the factors have been discarded:
  - Factor 4: Subject preference.
  - Factor 8: Prepositional phrase penalization.
- The five factors that have been left in the system, with some reservations:
  - Factor 5: Direct object preference.

- Factor 6: Indirect object preference.
- Factor 7: Adverbial phrase penalization.
- Factor 10: Section heading preference.
- Factor 11: Indefiniteness penalization.

Although these five factors either contributed little to the system, performed poorly, or even impaired the results, there is reason to believe that they are going to perform better in an improved version of ARN. The reason for this optimism is that they all have sound theoretical basis, and that there is good explanation for their poor performance: for Factor 10, it was too little data to perform on, and for the other four, there were problems associated with the pronoun *det* 'it (neut.)'. Their contribution could also be clearer with an evaluation module that could keep track of longer reference chains. I believe there is potential in them, and as long as they do not harm the system, they can be left to wait for a new and improved ARN.

The percent correctly resolved anaphora when using all factors was 71.0%. I have decided that the final version of ARN is going to contain all the factors except Factors 4 and 8, and this final version achieved 73.74% of correctly resolved anaphora when it was applied to the training corpus. It is this version of the system that was later been applied to the test corpus, and which is going to be analyzed in the next chapter.

Chapter 6

Results and evaluation

In this chapter I will present the final results of ARN.

The results were obtained by running ARN on a previously unknown set of data (test corpus). The evaluation of factors, which started in the previous section, will be continued. In the last part of this chapter, ARN's performance will be compared with two base-line models and the anaphora resolution systems we originally started with: RAP, MARS and Mitkov's original approach (MOA).

6.1 ARN's results on the test corpus

The results for the training and the test corpus are presented in tables 6.1 and 6.2, respectively. The results for the test corpus do not differ much from the results achieved with the training corpus: the overall result from the test corpus is only 3.23 percentage points lover than for the training corpus.

The distribution of results is more even within the training corpus, ranging from 66.09% to 83.62% against the results of the test corpus' files that range from 45.65% to 82.46%. File number 10 with its 45.65% has an exceptionally low percentage correctly resolved anaphora, but I have not managed to find anything in file 10 of the test corpus that would explain why so few anaphora are correctly resolved. The other files range from 63.33% to 82.46%, which is closer to file differences in the training corpus.

The highest percentage correctly solved anaphora (83.62%) was achieved on a file from the training corpus, but the second best result of 82.46% was achieved with a file from the test corpus. We will take a closer look at the factors' performance in both the test corpus and the training corpus later in this chapter, but from the numbers already presented here, we may conclude that ARN performed relatively well on an unknown set of data.

FILE NR.	Number of words	Number of anaphora	Correctly resolved	Wrongly resolved	No appropriate candidates	Percent correctly resolved
1	9916	115	76	38	1	66.08696
2	1969	116	97	16	3	83.62069
4	1988	129	92	34	3	71.31783
5	1995	29	23	4	2	79.31035
6	1984	71	50	21	0	70.42254
7	1970	90	72	16	2	80.0
8	1978	143	101	35	7	70.62937
TOTAL	21800	693	511	164	18	73.73737

Table 6.1: Applying ARN to the training data set

6.2 Evaluation of factors

In this section I will introduce two of Mitkov's (2002) measurements of factor performance: *Relative importance (RI)* and *Decision power (DP)*. I will start with the Relative importance because, as we will see, we have already been acquainted with this way of evaluating factors. The Decision power will hopefully cast some new light on them, in spite of some of the measurement's drawbacks.

6.2.1 Relative importance

Mitkov's *Relative importance* measure indicates a single factor's contribution to the system. Relative importance (RI_K) of a factor *K* is defined through equation (6.1).

$$RI_K = \frac{SR - SR_K}{SR} \tag{6.1}$$

where *SR* is the success rate of the system and SR_K is the success rate of the system when the factor *K* is excluded.

In other words, RI_K is a measurement of how much the system loses in performance when the factor *K* is removed.

If we look at the numbers for the training corpus in table 6.3, we can see that ARN's performance falls substantially when Factor 1 is removed; that Factor 7 neither contributes nor harms the system, while the presence of Factor 6 actually impairs the systems' performance.[1]

[1]In chapter 5 I introduced the measure "Criterion 2" which showed how many percent of the correctly resolved anaphora the system did loose when one factor was removed. Mitkov's *RI* is a

FILE NR.	Number of words	Number of anaphora	Correctly resolved	Wrongly resolved	No appropriate candidates	Percent correctly resolved
1	9290	91	67	23	1	73.62637
10	1988	92	42	39	11	45.65213
11	1993	70	51	18	1	72.85714
12	1976	114	83	29	2	72.80701
13	1998	60	38	18	4	63.33333
14	1976	171	141	29	1	82.45614
15	1978	134	85	44	5	63.43283
16	1966	122	94	27	1	77.04918
17	1988	85	61	24	0	71.76471
TOTAL	25153	939	662	251	26	70.50053

Table 6.2: Applying ARN to the test data set

When ARN was applied to the test corpus, **Factor 1: Number/Gender/Animacy factor** remained the one with most influence on the system, although its *RI* got lower in the test corpus (from 0.172 to 0.139).

Factor 2: The sentence proximity factor and **Factor 3: Boost pronoun** retained their position as the second and third most important factors, with almost the same increase in the test corpus, 0.023 and 0.022, respectively.

The **Factor 9: Syntactic parallelism**, which held a solid fourth position in the training corpus, performed slightly worse in the test corpus, and lost the fourth position to **Factor 11: Indefiniteness penalization**. RI_{11} increased more than 4 times, from the initially quite moderate 0.004 to 0.018. If we remember that Factor 11 belonged to the group of factors that were included into ARN "with some reservations", this was a pleasant surprise.

The *RI* of the rest of the factors that were included in ARN "with some reservations", namely factors **Factor 5: Direct object preference**, **Factor 6: Indirect object preference**, **Factor 7: Adverbial phrase penalization** and **Factor 10: Section heading preference**, remained mostly unaltered, varying by only $+/-0.002$. In the test corpus none of the factors impaired ARN's results. So, the five factors that were included in ARN on the basis of my trust in them did well, one of them surprisingly well, but it is of course not given that they would do that on any new data set.

In conclusion, we can say that the factors' influence on the system did not vary unaccountably between the known and the unknown set of data. Some changes

formalisation of that measurement and corresponds to the value of Criterion 2 divided by *SR*. The values of Criterion 2 for the training corpus are shown in the table in section 5.2 on page 73.

K	RI_K in train.	RI_K in test.
Factor 1: Number/Gender/Animacy factor	0.172	0.139
Factor 2: The sentence proximity factor	0.060	0.083
Factor 3: Boost pronoun	0.045	0.066
Factor 5: Direct object preference	0.004	0.003
Factor 6: Indirect object preference	-0.001	0.003
Factor 7: Adverbial phrase penalization	0.000	0.003
Factor 9: Syntactic parallelism	0.016	0.012
Factor 10: Section heading preference	0.002	0.000
Factor 11: Indefiniteness penalization	0.004	0.018
SR for the training corpus	73.74	
SR for the test corpus	70.05	

Table 6.3: The factors' relative importance (*RI*)

did occur, and it would have been be interesting to see why, but an investigation would severly break the time frame of this thesis.

6.2.2 Decision power

Mitkov (2002) defines the Decision power (DP_K) of a factor that awards positive points as the ratio between the number of candidates that the factor has awarded points to that got chosen as antecedent (S_K) and the total number of candidates that are awarded points by that factor (A_K):

$$DP_K = \frac{S_K}{A_K} \tag{6.2}$$

Decision power measures a single factor's impact without paying attention to its influence on the system (the factor's *RI*). For instance, a factor that only once awards points to a candidate gets the maximal *DP* of 1 if that candidate gets chosen. Other factors award points often because they are based on a quality that all candidates have, such as proximity to the anaphor. Giving points to many candidates ensures a high S_K value, so that the factor is almost certain to get a low *DP*, although it can many times be decisive for choosing the correct candidates.

For factors that give negative points (penalizing indicators), Mitkov (2002) defines this measure as:

$$DP_K = \frac{NonS_K}{A_K} \tag{6.3}$$

where $NonS_K$ is the number of cases where the candidate to which the indicator K has been applied has *not* been selected as the antecedent; A_K is again the total number of candidates penalized by the factor.

The *DP* definition for impeding factors is, in my opinion, the weakness of this measurement as it makes the measurement biased in favor of impeding factors. In the test corpus of ARN, there are over 12000 candidates, and only 939 anaphora, meaning that all the candidates have much higher chances not to be chosen as antecedents than to be chosen. For this reason, the two impeding factors of ARN, Factor 7 and Factor 11, and the impeding part of Factor 1[2] get seemingly undeservingly higher *DP*s than the rest of the factors. If we look at table 6.4, we can see that factor 11 penalized 3398 candidates; out of those 3398, only 60 were chosen as antecedents (S_K), which makes 3338 candidates not chosen, i.e. $NonS_K$ = 3338, so the DP_{11} ended up on 0.982. This number is so high in comparison to the other *DP*s that it can only be compared to *DP*s of the other impeding factors. The factor with the most decision power is the impeding part of Factor 1, but the *DP* values of other two impeding factors are very close to DP_1.

K	A_K	S_K	$NonS_K$	DP_K
Factor 1: Number/Gender/Animacy factor$_+$	3816	733	–	0.192
Factor 1: Number/Gender/Animacy factor$_-$	6552	30	6522	0.995
Factor 2: The reference proximity factor	8763	808	–	0.092
Factor 3: Boost pronoun	1449	580	–	0.400
Factor 5: Direct object preference	1447	82	–	0.057
Factor 6: Indirect object preference	57	7	–	0.123
Factor 7: Adverbial phrase penalization	461	6	455	0.987
Factor 9: Syntactic parallelism	4963	589	–	0.119
Factor 10: Section heading preference	94	14	–	0.149
Factor 11: Indefiniteness penalization	3398	60	3338	0.982

Table 6.4: The factors' decision power (*DP*), test corpus

Among the factors that give positive points, the factor with the by far best *DP* is Factor 3, with DP_3 = 0.4, followed by the boosting part of Factor 1, with the DP_1 of 0.192. Factors 9 and 2, which also constitute what I on page 77 called the "hard core" of ARN, did not get high *DP*s. Factor 9, with DP_9 = 0.119, is fifth, while Factor 2, with DP_2 = 0.092 ends up as next to the last of the boosting factors. This does not come as a surprise, as these factors are exactly the type that could be expected to have low *DP*s, as I remarked when I introduced the definition of

[2]As Factor 1 awards both positive and negative points to the candidates (see section 5.3.1, page 47), the only way to compute its *DP* was to split it into two factors. In tables 6.4 and 6.5 they are marked with $_+$ and - indexes.

Decision power. The good rating of Factor 10, which has the third best *DP*, is not surprising either. When introducing Factor 10 in section 5.3.6, I said that although it brought only one correct resolution, it did exactly what it was expected to do, helping the candidate which appeared in the section title to the first place of the candidate list. I left the factor in ARN believing that it did a good job, and that its weak performance was due to the lack of opportunities to be used. Its relatively high *DP* reflects this.

I also computed the *DP*s for the training corpus, but apart from factors 6 and 9 swapping positions, I did not find any significant changes in the factors' *DP*s when they were applied to the training corpus and the test corpus. The results for both corpora are given in the table 6.5.

K	DP_K training	DP_K test
Factor 1: Number/Gender/Animacy factor$_+$	0.215	0.192
Factor 1: Number/Gender/Animacy factor$_-$	0.995	0.995
Factor 2: The reference proximity factor	0.097	0.092
Factor 3: Boost pronoun	0.414	0.400
Factor 5: Direct object preference	0.050	0.057
Factor 6: Indirect object preference	0.109	0.123
Factor 7: Adverbial phrase penalization	0.992	0.987
Factor 9: Syntactic parallelism	0.138	0.119
Factor 10: Section heading preference	0.180	0.149
Factor 11: Indefiniteness penalization	0.988	0.982

Table 6.5: The factors' decision power (*DP*)

The analysis of both the *RI* and *DP* measurements shows that the factors' performance in the system is a complex question, and that much more data and deeper analysis is needed to throw more light on it.

6.3 Comparison with other systems

6.3.1 Comparison with baseline models

ARN has been compared with the following baseline models:

- B1: The closest candidate (i.e. the closest noun) is proposed as antecedent.
- B2: The closest candidate that matches in gender and number is proposed as antecedent.

When baseline model B1 was applied to the test set of data, it resolved correctly 22.36% of the anaphora, while model B2 achieved a result of 45.56%. Both of these results are thus considerably lower than ARNs 70.50%.

6.3.2 Comparison with other AR systems

I will start this section by following Mitkov's proposal to distinguish between comparing anaphora resolution *systems* and anaphora resolution *algorithms* (Mitkov 2002).

Only a comparison of anaphora resolution algorithms could give us fair insight into the systems' relative performance. However, comparing anaphora resolution algorithms would imply implementing all the algorithms in question and letting them run in an 'ideal environment', meaning that all the algorithms would have the same set of manually checked data, and that the output of all the systems we compare would be manually checked and analyzed. That is, of course, a project of its own, and certainly not something that is possible to do at the final stage of a hovedfag thesis.

We may compare reported results, but as it turns out, they will not give us a satisfactory comparison. The key numbers (percent of correctly resolved anaphora) for ARN and the three systems it was inspired by are the following:

- ARN: 70.50%
- RAP: 84.00%
- MOA: 89.70%
- MARS: 61.55%

There are several reasons why these anaphora resolution systems are difficult to compare, and they can be grouped into three main classes:

1. Differences in aims
2. Differences in pre-processing
3. Differences in scope (genre)
4. Differences in resources

Differences in aims. All the systems we are comparing ARN with have been designed to resolve all personal pronouns, while ARN's aim has been to resolve the third person pronouns apart from the pronoun *det* 'it (neut.)'. Even so, ARN can in principle handle all personal pronouns and in that case, the percentage of correctly resolved anaphora is 62.61%. However, this number cannot be compared straight-forwardly with the performance of the other systems since ARN was not intended to be able to resolve all personal pronouns – the system does for instance not take into consideration dialogues, coordination, and other data essential for resolving the first and second person pronouns.

Differences in pre-processing. As Mitkov (2001) and others have noticed, inaccurate pre-processing leads to a drop in the performance of anaphora resolution systems no matter how accurate the algorithm may be. The mistakes that stem from pre-processing tend to propagate through the system and considerably affect the final results. The systems we are comparing are very different from each other with respect to pre-processing.

The systems with the highest results (RAP and MOA) had their data manually checked and corrected before it was send to the anaphora resolution module, preventing in that way the pre-processing mistakes to multiply through the system. When it, for instance, comes to the problem of ambiguity of the anaphor *it* with the expletive subject *it*, in RAP it was resolved by simply manually removing expletives.

On the other hand, MARS, which performed worst, did everything without any human intervention, from accepting a raw text to producing an output with the anaphora resolved.

In this aspect, ARN comes somewhere in between. It used input data automatically tagged by the Oslo-Bergen tagger. However, in some cases of ambiguity, the manually tagged data from the BREDT corpus was used for disambiguation.

Differences in genre. Both MARS and MOA were trained and tested on technical manuals, so several of the antecedent indicators were clearly genre-specific. RAP was tested on computer manual texts. ARN has been trained and tested on newspaper articles and literary texts ranging from classical literature over love novels of dubious quality to folk legends, which excluded the possibility of boosting the system with genre-specific factors.

Differences in resources. And, finally, ARN was made by a single post-graduate student with all the limitation in experience, time and resources that position implies.

6.4 Summary

In this chapter I have presented the results that ARN has achieved when given an unknown set of data. On the test corpus it resolved correctly 70.50% of anaphora, or 662 out of 939. The resolution factors were evaluated according to their importance for the system as whole, Relative importance (*RI*), and how big impact a single factor has for promoting the correct candidate, the Decision power (*DP*).

The final version of ARN was compared to two baseline models, one that chose the closest noun as a candidate, and one that chose the closest noun matching in gender and number. Finally, ARN was compared to MARS, MOA and RAP, the three systems it was initially inspired by, only to conclude that this comparison is quite unreliable, due to differences in aims, pre-processing, genre and resources.

Chapter 7

Conclusion

In this chapter I will sum up the achievements of ARN and give a brief outline of possible improvements of the system.

7.1 What has been accomplished?

In this thesis I have presented *ARN – an Automatic Anaphora Resolution System for Norwegian.* The factors that were initially planned to constitute the anaphora resolution module of ARN are:

1. Number/Gender/Animacy factor
2. The sentence proximity factor
3. Boost pronoun
4. Subject preference
5. Direct object preference
6. Indirect object preference
7. Adverbial phrase penalization
8. Prepositional phrase penalization
9. Syntactic parallelism
10. Section heading preference
11. Indefiniteness penalization

However, during experimentation with the factors, I found out that a substantial group of them were not as successful when applied to Norwegian as they were for English, some of them being actually directly counterproductive. These factors have in common that they are based on the the following super-rule:

> Prefer candidates that are subjects to candidates that are direct objects, prefer direct objects to indirect objects, and prefer indirect objects to other constituents such as adverbial or prepositional phrases.

The factors involved were:

4. Subject preference
5. Direct object preference
6. Indirect object preference
7. Adverbial phrase penalization
8. Prepositional phrase penalization

and, to a certain extent,

11. Indefiniteness penalization

This was a surprising find considering that these rules constitute the majority of rules in the RAP system and otherwise have a solid base in the English-oriented systems.

I have come to the conclusion that this is due to differences in information structure between Norwegian and English. The stronger reluctance towards conveying new information by subject leads to a greater occurrence of expletive subjects in Norwegian, leaving them less likely to be antecedents. This difference in information structure does not only concern candidates in subject position, but also influences salience of indirect and direct objects and adverbial and prepositional phrases. This eventually led to the exclusion of two factors from the system: Factor 4 (Subject preference) and Factor 8 (Prepositional phrase penalization). The rest of the factors in this group were left in the system, although runs on the training data did not give this decision unambiguous support. When ARN was applied to a new data set these factors performed better, which gave me reason to believe that a bigger and more diversified data would see the better expression of these rules.

This finding shows that there are differences between English and Norwegian that are important enough to create a need for an anaphora resolution system that takes the structure of Norwegian into account. ARN has done just that, achieving a solid result of 70.50% correctly resolved anaphora.

7.2 Future prospects

I have mentioned numerous times in this thesis the need for prolonging reference chains and finding a way to identify expletive *det* 'it'. Apart from those, there is a number of possibilities for further development of ARN. To begin with, ARN could be provided with an NP-chunker and a user-interface connected to the Oslo-Bergen tagger. It would be interesting to see how ARN would perform on a corpus that has dialog markings, which would make it possible to look at the first and second person pronouns as well. The resolving anaphora *we* and *you* brings the question of solving coordination of NPs, which would probably also improve the results of resolving the anaphor *de* 'they'. An improved evaluation of factors could also lead to a better system.

And finally, there is the interesting possibility of combining the rule-based and statistical approaches and supplement ARN with a statistical optimization of its factors.

Appendix A

The programming code

ARN has been implemented in LISP. Its programming code is too long to be presented here in its entirety, so I have chosen to present only parts of it in this Appendix. The complete code with all the data files is enclosed on the CD that follows with the thesis.

I will start with one of the top procedures, the procedure `fullan`. Its output, as well as examples of the input files are given in the next part of the Appendix, in section B.2.4 (p.117) and section B.1 (p.101), respectively.

The `fullan` procedure takes as argument a number that represents file number. I will show only the beginning of the procedure `fullan`, as the remainder of it opens other files in the same manner:

```
(defun fullan (x)
  (let
     ((path nil)
      (stream nil)
      (path2 nil)
      (stream2 nil))
    (setf
        *BUFFER* ""
        *NEW-WORD* nil
        *NEW-SENTENCE* nil
        *S-TOKENS* nil
        *S-PREVIOUS-1* nil
        (...)
        *STACK2* nil)
    (case x
      (1
       (setf
           path (make-pathname :name "file1-oslo.txt")
           stream (open path :direction :input)
           path2 (make-pathname :name "file1-bredt.txt")
           stream2 (open path2 :direction :input))
```

```
      (make-sentence-objects-2streams stream stream2)[1]
      (close stream)
      (close stream2)
      (analysis))
    (10
      (setf
          path (make-pathname :name "file10-oslo.txt")
(...)
```

The procedure `fullan` opens two versions of the file - the Oslo and the BREDT version, as presented in section 4.1, on page 30. It calls subprocedure `make-sentence-objects-2streams` that reads through the file, making word- and sentence-objects. When a sentence object is made, `make-sentence-objects-2streams` checks it for the presence of anaphora, and if anaphora are found, it performs anaphora resolution. After the whole file is processed, `fullan` invokes the subprocedure `analysis` that prints out the statistics on the anaphora resolution for the whole file.

Most of the work in connection with anaphora resolution is done by the `make-sentence-objects-2streams` procedure. This function takes as input file streams (opened by `fullan`):

```
(defun make-sentence-objects-2streams (stream stream2)
 (let ((L '())
       (sent (make-instance 'sentence))
       (S (read-sentence-2streams stream stream2)))
   (do ((teller 0 (+ teller 1)))
       ((or *NEW-DOCUMENT* (equal S 'EOF)
            (equal *TITLEBUFFER* '(eof eof)))
        (if *NEW-DOCUMENT*
            (progn
             (setf (s-previous-2 sent) nil
                   (s-previous-1 sent) nil
                   (s-tokens sent) S)
            (resolve sent)

            (display-list (s-tokens sent))
            (format t "~%")
            (push sent *LST*)
            (setf
                *S-TOKENS* nil
                *S-PREVIOUS-1* nil)
            (setf *NEW-DOCUMENT* nil
                  *TITLE* nil)
            (setf L (make-sentence-objects-2streams stream stream2)))
         *LST*))
   (setf (s-previous-2 sent) *S-PREVIOUS-1*
```

[1]The subprocedures that we will take a closer look at are underlined.

```
            (s-previous-1 sent) *S-TOKENS*
            (s-tokens sent) S)

   (resolve sent)
      (display-list (s-tokens sent))
      (format t "~%")
      (push sent *LST*)
      (setf *S-TOKENS* (s-tokens sent)
            *S-PREVIOUS-1* (s-previous-1 sent))
      (setf sent (make-instance 'sentence))
      (setf S (read-sentence-2streams stream stream2)))
     (cons L *LST*))
    (reverse *LST*))
```

The procedure `make-sentence-objects-2streams` starts by making an empty sentence-object by invoking a build-in procedure `make-instance`. It then calls the subprocedure `read-sentence-2streams`. This procedure makes word-objects and puts them to a list that corresponds to one sentence in the file. This list is put into the `s-tokens`-slot of the new-made sentence-object. Upon reading new sentences, `make-sentence-objects-2streams` will move the list from the `s-tokens` slot first to the `s-previous-1` and then `s-previous-2` slots of new sentence-objects, while the emptied slots will be filled with new sentences. In that way, for each new sentence in the file `make-sentence-objects-2streams` makes a three-sentence window where anaphora resolution is performed. The procedure `make-sentence-objects-2streams` starts the process of anaphora resolution by invoking the subprocedure `resolve`:

```
(defun resolve (sobj)
  (let ((pronlst (reverse (pronounlst (s-tokens sobj)))))
    (if (null pronlst)
        (progn
          (display-lemmas-2 (s-tokens sobj))
          (format t "There are no anaphors in this sentence.
~%**************************************************~%")
          (setf *NO-ANAPHORS* (+ *NO-ANAPHORS* 1)))
       (progn
        (display-all-lemmas sobj)
        (format t "The considered anaphors in this sentence are:
~{ ~S~}.~%~%" (clear-wt-lst pronlst))
        (substack-so sobj)
        (setf *STACKS* (stacks))

       (factors pronlst (stacks) sobj))))
       (setf *STACK0* nil
```

```
            *STACK1* nil
            *STACK2* nil
            *STACKS* nil))
```

The procedure `resolve` starts by extracting the pronouns from the sentence-object's `s-tokens` slot. It is done by the `pronounlst` subprocedure. If there are anaphora to be resolved, it lists them and makes a stack of candidates from the current sentence object by invoking the procedure `substack-so`. When the lists of anaphora and their resolution candidates are ready, `resolve` invokes the `factors` procedure:

```
(defun factors (pronlst stacklst sobj)
  (if (null pronlst)
      (format t "Resolution of the sentence object completed
~%**************************************************~%~%")
    (let ((pron (car pronlst)))
      (factor1-14 pron stacklst)
      (format t "Factor 1 (~A) applied.~%" (factors-id 1))

      (factor2 pron stacklst)
      (format t "Factor 2 (~A) applied.~%" (factors-id 2))
      (factor5 pron stacklst)
      (format t "Factor 5 (~A) applied.~%" (factors-id 5))
(...)
   (factor11 pron stacklst)
      (format t "Factor 11 (~A) applied.~%" (factors-id 11))
      (display-resolution-fact pron *STACKS* sobj)
      (setf *STACKS* nil
            *STACK0* nil
            *STACK1* nil
            *STACK2* nil)
      (substack-so sobj)
      (setf *STACKS* (stacks))
      (factors (cdr pronlst) *STACKS* sobj)
  )))
```

The procedure `factors` takes as input list of pronouns to be resolved, list of the resolution candidates and the sentence-object that is the scope of resolution. The resolution starts by applying each factor on each candidate. If there is more than one anaphor in the list, `factors` makes a new candidate stack, and begins the process anew, until the list of pronouns is empty. I will here show the `factor2` procedure that is representative for the rest of the `factor`*n* procedures, with the exception of `factor1` that is considerably more complex, though built on the same principle.

```
(defun factor2 (pron stacklst)
  (if (null stacklst)
      nil
```

```
    (let*
((prvi (car stacklst))
 (cand (cadar stacklst))
 (pid  (wt-id pron))
 (snrp (wt-sent-nr pron))
 (snrc (wt-sent-nr cand)))
      (cond
       ((= snrp snrc)
(set-points prvi 100 2 pid)
(setf (gethash 'ak (gethash 2 decpower))
      (+ 1 (gethash 'ak  (gethash   2 decpower)))))
       ((= snrp (+ snrc 1))
(set-points prvi 50  2 pid)
(setf (gethash 'ak (gethash   2 decpower))
      (+ 1 (gethash 'ak (gethash   2 decpower)))))
       ((= snrp (+ snrc 2))
(set-points prvi 0   2 pid)))
      (setf *STACKS* (cons prvi (factor2 pron (cdr stacklst))))
      )))
```

All the `factorn` procedures take the anaphor to be resolved and a list of candidates as input. The procedure `factor2` compares the sentence numbers of the candidate (`snrc`) and the anaphor (`snrp`), giving the candidates points that correspond to their proximity to the anaphor. When factors award point sums $\neq 0$, it is registered in the `decporwer` hash-table, and this data is later used for evaluation of factors. By the end of each resolution `factors` invoke `display-resolution-fact`:

```
(defun display-resolution-fact (pron stacklst sobj)
  (cond ((and pron  (null stacklst))
 (format t "There are no candidates in the three sentence scope.
~%**************************************************~%")
    (setf *NO-CANDIDATES* (+ *NO-CANDIDATES* 1)))
((null stacklst)
 (format t "There are no pronominal anaphors in this sentence. ~%~%")
 (format t
"**************************************************~%"))
        (t
         (let*
             ((pickstack  (pick-out pron  stacklst))
      (newstack (sort pickstack #'> :key #'car)))
           (if (null newstack)
               (progn
                 (format t "There are no candidates in
 the three sentence scope.~%")
                 (setf *NO-CANDIDATES* (+ *NO-CANDIDATES* 1)))
             (progn
               (setf (aref (wt-extra-features pron) 3)
```

```
                    (wt-id (cadar newstack)))
                (factor3 pron newstack)
                (format t "Factor 3 (~A) applied.~%" (factors-id 3))
                (setf newstack  (sort newstack #'> :key #'car))
              (if (< (caar newstack) *THRESHOLD*
                  (progn
                    (display-stack-single newstack pron)
                    (format t "There are no appropriate candidates
in the three sentence scope, as none of the
candidates reaches over the threshold  of 0~%")
                    (setf *NO-CANDIDATES* (+ *NO-CANDIDATES* 1)))
                (progn
                   (format t "The anaphor ~:@(~A~) has
~(~R~) resolution candidate~:P.
The numbers in front are their point scores.~%~%"
                           (string-trim "<>"(wt-word pron))
                           (length newstack))
                   (setf (wt-ant pron) (cadar newstack)
                   (display-stack-single newstack pron)
                   (format t "The proposed referent of anaphor ~:
@(~A~) is ~:@(~A~)~%"
                           (string-trim "<>"(wt-word pron))
                           (string-trim "<>"(wt-word (cadar newstack))))

                   (check-if-correct pron (cadar newstack))


                   (format t
"------------------------------------------------------------~%~%")
                   )))))))
  (setf *STACKS* nil
        *STACK0* nil
        *STACK1* nil
        *STACK2* nil)
  (substack-so sobj)
  (setf *STACKS* (stacks)))
```

The name `display-resolution-fact` is somewhat misleading, for besides displaying results, this procedure performs several other important chores: It applies the `factor3` procedure that has to be applied after all the other factors; it keeps count of the cases where no appropriate candidate was found within the three-sentence scope, and it checks if the resolution has been successful by invoking `check-if-correct`:

```
(defun check-if-correct (pronoun noun)
  (format t "The reference of pronoun ~A is ~A~%"
          (wt-word  pronoun) (wt-word noun))
  (when (wt-ant pronoun)
    (when (wt-ref (wt-ant pronoun))
```

```
      (setf (wt-ref pronoun)
            (append
             (wt-ref (wt-ant pronoun)) (wt-ref pronoun)))))
  (if (member (wt-id noun) (wt-ref pronoun))
       (progn (setf *CORRECT* (+ *CORRECT* 1))
     (format t "(CORRECT!)~%~%"))
    (progn (setf *WRONG* (+ *WRONG* 1))
   (format t "(WRONG!)~%~%")))
  T)
```

The procedure `check-if-correct` evaluates the resolution. If an anaphor has another anaphor as its antecedent, and they are annotated in such a way that they have identical references, the reference numbers of the candidate are appended to the reference numbers of the anaphor. In practice, this means that if anaphor a_1 refers to anaphor a_2, and a_2 refers to noun n_1, then if the anaphor a_1 is resolved to the noun n_1, this resolution is rendered correct. If the candidate is a noun, its identification number has to be contained in the anaphor's reference number to be identified as a successful resolution. It is done to cover for the cases where the anaphor refers to a noun phrase, but only the head of the phrase is picked up as the antecedent. This case also covers the simplest kind of correct resolution, namely when the id-number of the candidate is identical to the anaphor's reference number.

The procedure `check-if-correct` also keeps count on the total number of correct and wrong resolutions through the global parameters `*CORRECT*` and `*WRONG*`. Back in the top `fullan` procedure, this data is used by the `analysis` procedure:

```
(defun analysis ()
  (setf *TOTAL* (- *TOTAL* 1))
   (format t
"~%-------------------------------------------------~%
In ~A sentences, there are ~A pronominal anaphors:
~A sentences contain one or more anaphors,
~A sentences do not contain any.~%
Number of correctly resolved anaphors: ~A~%
Number of wrongly resolved anaphors: ~A~%
Number of anaphors where no appropriate candidate
was found in a three sentences scope: ~A~%
-------------------------------------------------~%"
  *TOTAL*
  (+   *CORRECT*
       *WRONG*
       *NO-CANDIDATES*)
  (- *TOTAL*
     *NO-ANAPHORS*)
  *NO-ANAPHORS*
```

```
*CORRECT*
*WRONG*
*NO-CANDIDATES*))
```

Now that data on resolution is gathered from the whole file, `analysis` prints the data on the total number of resolutions, the number of correct and wrong resolutions and the cases where no appropriate candidate was found, and finishes in this way the processing of the file.

Appendix B

Input and output

B.1 The data

The file used in these examples is file 11 of the test corpus, chosen because ARNs success rate on this file (72.86%) is not far from ARN's overall success rate on the test corpus (70.50%). Like all other data files used by ARN, the file 11 consists of two parallel files. The first sentence of both of the files is shown below.

B.1.1 Oslo file

```
"<Johan Oles>"
        "Johan Oles" subst prop mask gen @det> &person &sted &org &verk &hend &annet rx
"<matematikklrer>"
        "matematikklrer" subst appell mask ub ent @subj
"<holdt>"
        "holde" verb pret tr1 tr11 pa1 a9 pr7 pr6 @fv
"<paa>"
        "paa" prep @adv
"<aa>"
        "aa" inf-merke @<p-utfyll
"<gjennomgaa>"
        "gjennomgaa" verb inf tr1 tr2 @iv
"<en>"
        "en" det mask ent kvant @det>
"<oppgave>"
        "oppgave" subst appell mask ub ent @obj
"<.>"
        "$." clb <<< <punkt>
"<Gutten>"
        "gutt" subst appell mask be ent @subj
"<burde>"
        "burde" verb pret <aux1/infinitiv> @fv
"<altsaa>"
        "altsaa" adv @adv
"<tenke>"
        "tenke" verb inf tr1 i1 pa1 tr2 tr12 tr3 rl8 rl7 @iv
"<paa>"
        "paa" prep @adv
```

```
"<matematikk>"
        "matematikk" subst appell mask ub ent @<p-utfyll
"<.>"
        "$." clb <<< <punkt>
```

B.1.2 BREDT file

```
b( '0', 'Johan Oles', 'les', 'Ole', 'prpg', 'det' ).
b( '2', 'matematikklrer', 'rer', 'matematikklrer', 'ni', 'subj' ).
b( '3', 'holdt', 'ldt', 'holde', 'v', 'fv' ).
b( '4', 'paa', 'paa', 'paa', 'p', 'adv' ).
b( '5', 'aa', 'aa', 'aa', 'inf', 'pfill' ).
b( '6', 'gjennomgaa', 'mgaa', 'gjennomgaa', 'v', 'iv' ).
b( '7', 'en', 'en', 'en', 'd', 'det' ).
b( '8', 'oppgave', 'ave', 'oppgave', 'ni', 'obj' ).
b( '9', '.', '.', '$.', 'dot', 'unk' ).
b( '10', 'Gutten', 'ten', 'gutt', 'nd', 'subj', '0' ).
b( '11', 'burde', 'rde', 'burde', 'v', 'fv' ).
b( '12', 'altsaa', 'tsaa', 'altsaa', 'adv', 'adv' ).
b( '13', 'tenke', 'nke', 'tenke', 'v', 'iv' ).
b( '14', 'paa', 'paa', 'paa', 'p', 'adv' ).
b( '15', 'matematikk', 'ikk', 'matematikk', 'ni', 'pfill' ).
b( '16', '.', '.', '$.', 'dot', 'unk' ).
```

B.2 Examples of program output

ARN has four top procedures. They all perform anaphora resolution in the same way, but present the results on different detail levels. To get an overview of the results from the resolution of all the anaphora of the test corpus, one can use the procedure `table` (p. 102). The procedure `result` (p. 103) is used to get the corresponding results for a single file. To get insight into the sentence objects and for a detailed account of each single resolution in a file, ARN has procedure `partan` (p. 103). Finally, the most detailed procedure, `fullan` (p. 117) shows all word-objects, all sentence-objects and a detailed account of each resolution.

B.2.1 Table

The call on a command `table` produced the data that are used for table 6.2 showed on page 81. In the LISP version, the table looks like this:

```
CL-USER(2): (table)
```

|file	|# sent	|# w/an	|# wo/an	|# anaph	|correct	|wrong	|no cands|	% corr.
|FILE 1	|589	|78	|511	|91	|67	|23	|1	|73.62637
|FILE 10	|175	|73	|102	|92	|42	|39	|11	|45.652172
|FILE 11	|171	|56	|115	|70	|51	|18	|1	|72.85714

|FILE 1	|112	|63	|49	|114	|83	|29	|2	|72.807014
|FILE 13	|193	|45	|148	|60	|38	|18	|4	|63.333332
|FILE 14	|130	|102	|28	|171	|141	|29	|1	|82.45614
|FILE 15	|109	|81	|28	|134	|85	|44	|5	|63.432835
|FILE 16	|102	|62	|40	|122	|94	|27	|1	|77.04918
|FILE 17	|106	|52	|54	|85	|61	|24	|0	|71.76471
TOTAL	|1687	|612	|1075	|939	|662	|251	|26	|70.500534

```
Using all factors (except factors 4 and 8)
on all pronouns
Percent correctly resolved: 70.500534
```

B.2.2 Results only

This output puts essentially one row of the table in words. In our case, it is the row concerning file 11:

```
CL-USER(3): (result 11)
FILE 11:

------------------------------------------------------------

In 171 sentences, there are 70 pronominal anaphors: 56 sentences contain one or more anaphors,
115 sentences do not contain any.

Number of correctly resolved anaphors: 51

Number of wrongly resolved anaphors: 18

Number of anaphors where no appropriate candidate was found in a three sentences scope: 1
------------------------------------------------------------
NIL
```

B.2.3 Partial analysis

The `partan` procedure shows the resolution of the text on the sentence-object level. Although file 11 is of average length (171 sentences), the print-out of the anaphora resolution performed on it took 70 pages of text, in the small font presented here. The next and more detailed level of analysis shown in section B.2.4 (p. 117) of this appendix, is longer than 500 pages. Since I cannot show complete analysis here, I will show only the first ten or so pages of each.

```
CL-USER(5): (partan 11)
Johan Oles matematikklaerer holdt paa aa gjennomgaa en oppgave .
There are no anaphors in this sentence.
**********************************************************
```

```
Gutten burde altsaa tenke paa matematikk .
There are no anaphors in this sentence.
************************************************************
S-tokens: Men det gjorde han ikke .

S-previous-1: Gutten burde altsaa tenke paa matematikk .

S-previous-2: Johan Oles matematikklaerer holdt paa aa gjennomgaa en oppgave .


The considered anaphors in this sentence are:  "han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has five resolution candidates.
The numbers in front are their point scores.

150.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
50.0 Johan Oles (id-nr 0)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-37.5 oppgave (id-nr 8)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-37.5 matematikklaerer (id-nr 2)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-56.25 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
The proposed referent of anaphor HAN is GUTTEN
(wt-id reference): 10
(wt-features reference): ((gutt) (subst appell mask be ent @subj))
(wt-extra-features reference): #(1 0 3 -1)
(wt-id anaphor): 20
(wt-ref anaphor): (10)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(4 0 3 10)
(CORRECT!)

---------------------------------------------------------------------

Resolution of the sentence object completed
*********************************************************************

S-tokens: Hvis noen hadde spurt ham hva han tenkte paa , ville han ikke kunnet svare .
```

```
S-previous-1: Men det gjorde han ikke .

S-previous-2: Gutten burde altsaa tenke paa matematikk .


The considered anaphors in this sentence are:  "ham" "han" "han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAM has three resolution candidates.
The numbers in front are their point scores.

187.5 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
50.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAM is HAN
(wt-id reference): 20
(wt-features reference): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features reference): #(4 0 3 10)
(wt-id anaphor): 27
(wt-ref anaphor): (20)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 akk @obj))
(wt-extra-features anaphor): #(5 0 3 20)
(CORRECT!)

---------------------------------------------------------------------


Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has four resolution candidates.
The numbers in front are their point scores.

325 ham (id-nr 27)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
225.0 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
```

```
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
75.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAN is HAM
(wt-id reference): 27
(wt-features reference): ((han) (pron mask ent pers hum 3 akk @obj))
(wt-extra-features reference): #(5 0 3 20)
(wt-id anaphor): 29
(wt-ref anaphor): (27)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(7 0 3 27)
(CORRECT!)

---------------------------------------------------------------------

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has five resolution candidates.
The numbers in front are their point scores.

325 han (id-nr 29)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
325 ham (id-nr 27)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
225.0 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
75.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAN is HAN
(wt-id reference): 29
(wt-features reference): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features reference): #(7 0 3 27)
(wt-id anaphor): 34
(wt-ref anaphor): (29)
```

```
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(12 0 3 29)
(CORRECT!)

----------------------------------------------------------------

Resolution of the sentence object completed
********************************************************************************

S-tokens: Han ville sagt : Paa ingenting .

S-previous-1: Hvis noen hadde spurt ham hva han tenkte paa , ville han ikke kunnet svare .

S-previous-2: Men det gjorde han ikke .


The considered anaphors in this sentence are:  "Han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has four resolution candidates.
The numbers in front are their point scores.

225.0 han (id-nr 34)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
225.0 han (id-nr 29)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
225.0 ham (id-nr 27)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
150.0 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAN is HAN
(wt-id reference): 34
(wt-features reference): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features reference): #(12 0 3 29)
(wt-id anaphor): 39
(wt-ref anaphor): (34)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(1 0 3 34)
(CORRECT!)

----------------------------------------------------------------

Resolution of the sentence object completed
********************************************************************************
```

```
Og det ville vaert temmelig naer sannheten .
There are no anaphors in this sentence.
************************************************************
Men noen tanker var der : tanker saa skiftende som horisonten i vest , og saa vage som droemmer .
There are no anaphors in this sentence.
************************************************************
S-tokens: Han hadde arvet lange og litt tunge lemmer av sin bondeslekt .

S-previous-1: Men noen tanker var der : tanker saa skiftende som horisonten i vest ,
og saa vage som droemmer .

S-previous-2: Og det ville vaert temmelig naer sannheten .


The considered anaphors in this sentence are:  "Han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
-18.75 tanker (id-nr 60)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-18.75 tanker (id-nr 56)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-25.0 sannheten (id-nr 52)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-37.5 horisonten (id-nr 64)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
There are no appropriate candidate in the three
  sentence scope, as none of the candidates reaches over the threshold
  of 0
Resolution of the sentence object completed
************************************************************************

I oeyeblikket var kroppen fullstendig avslappet og liksom toet inn i alle krokene av pulten .
There are no anaphors in this sentence.
************************************************************
De blaagraa oeynene var fravaerende , den brede munnen litt loes .
There are no anaphors in this sentence.
************************************************************
Det lyse haaret var forsoekt glattet til side ,
men en bunt haarstraa sto ende til vaers .
There are no anaphors in this sentence.
************************************************************
S-tokens: Lektor Myrseths stemme gjaldt plutselig ham : - Ole !

S-previous-1: Det lyse haaret var forsoekt glattet til side ,
 men en bunt haarstraa sto ende til vaers .
```

```
S-previous-2: De blaagraa oeynene var fravaerende , den brede munnen litt loes .


The considered anaphors in this sentence are:  "ham".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAM has ten resolution candidates.
The numbers in front are their point scores.

200 Myrseths (id-nr 132)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
175 Lektor (id-nr 131)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
-25 stemme (id-nr 133)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
-37.5 haaret (id-nr 116)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-50.0 munnen (id-nr 110)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-50.0 oeynene (id-nr 104)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-56.25 haarstraa (id-nr 126)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-56.25 bunt (id-nr 125)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-56.25 side (id-nr 121)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-93.75 ende (id-nr 128)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, -50 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
The proposed referent of anaphor HAM is MYRSETHS
(wt-id reference): 132
(wt-features reference): ((Myrseth) (subst prop gen @det> &person <person>))
```

```
(wt-extra-features reference): #(2 0 1 -1)
(wt-id anaphor): 136
(wt-ref anaphor): (74)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 akk @obj))
(wt-extra-features anaphor): #(6 0 3 132)
(WRONG!)

------------------------------------------------------------------

Resolution of the sentence object completed
**************************************************************************

 Ikke svar .
There are no anaphors in this sentence.
**********************************************************
S-tokens: Johan Ole hadde hoert navnet , men det var ikke helt bevisst for ham hva det betydde .

S-previous-1:  Ikke svar .

S-previous-2: Lektor Myrseths stemme gjaldt plutselig ham : - Ole !


The considered anaphors in this sentence are:  "ham".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAM has seven resolution candidates.
The numbers in front are their point scores.

200 Johan Ole (id-nr 145)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
150.0 ham (id-nr 136)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
50.0 Ole (id-nr 139)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
50.0 Myrseths (id-nr 132)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
37.5 Lektor (id-nr 131)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 stemme (id-nr 133)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-75 navnet (id-nr 149)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
```

```
 0 pts by factor 9, -50 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
_____________________________________________________________________
The proposed referent of anaphor HAM is JOHAN OLE
(wt-id reference): 145
(wt-features reference): ((Johan Ole) (subst prop mask @subj &person &org rx))
(wt-extra-features reference): #(1 0 3 -1)
(wt-id anaphor): 158
(wt-ref anaphor): (145)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 akk @<p-utfyll))
(wt-extra-features anaphor): #(13 0 3 145)
(CORRECT!)

---------------------------------------------------------------------

Resolution of the sentence object completed
*********************************************************************

- Ole !
There are no anaphors in this sentence.
***********************************************************
 Naa klarte bevisstheten aa tolke hva stemmen sa .
There are no anaphors in this sentence.
***********************************************************
S-tokens: Han snudde seg mot kateteret og saa like inn i lektor Myrseths briller .

S-previous-1:  Naa klarte bevisstheten aa tolke hva stemmen sa .

S-previous-2: - Ole !


The considered anaphors in this sentence are:  "Han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has three resolution candidates.
The numbers in front are their point scores.

50.0 Ole (id-nr 164)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
0.0 stemmen (id-nr 173)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
0.0 bevisstheten (id-nr 169)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
The proposed referent of anaphor HAN is OLE
(wt-id reference): 164
(wt-features reference): ((Ole) (subst prop mask @loes-np &person <person>))
(wt-extra-features reference): #(2 0 3 -1)
```

```
(wt-id anaphor): 176
(wt-ref anaphor): (158)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(1 0 3 164)
(WRONG!)

------------------------------------------------------------------------

Resolution of the sentence object completed
************************************************************************

S-tokens: Glassene var tykke , og det blinket i dem .

S-previous-1: Han snudde seg mot kateteret og saa like inn i lektor Myrseths briller .

S-previous-2:  Naa klarte bevisstheten aa tolke hva stemmen sa .


The considered anaphors in this sentence are:  "dem".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor DEM has eight resolution candidates.
The numbers in front are their point scores.

100 Glassene (id-nr 190)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 0 pts by factor 1, multiplied by 1
56.25 briller (id-nr 188)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 0 pts by factor 1, multiplied by 0.75
0.0 kateteret (id-nr 180)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-37.5 Myrseths (id-nr 187)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-50 i (id-nr 197)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, -50 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
-50.0 stemmen (id-nr 173)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-50.0 bevisstheten (id-nr 169)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-56.25 lektor (id-nr 186)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
```

```
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
The proposed referent of anaphor DEM is GLASSENE
(wt-id reference): 190
(wt-features reference): ((glass) (subst appell noeyt be fl @subj))
(wt-extra-features reference): #(1 0 0 -1)
(wt-id anaphor): 198
(wt-ref anaphor): (190)
(wt-features anaphor): ((de) (pron fl pers 3 akk @<p-utfyll @obj))
(wt-extra-features anaphor): #(9 0 0 190)
(CORRECT!)

---------------------------------------------------------------------

Resolution of the sentence object completed
**************************************************************************

S-tokens: - De kan ta oppgave 21 .

S-previous-1: Glassene var tykke , og det blinket i dem .

S-previous-2: Han snudde seg mot kateteret og saa like inn i lektor Myrseths briller .


The considered anaphors in this sentence are:  "De".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor DE has seven resolution candidates.
The numbers in front are their point scores.

112.5 dem (id-nr 198)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 0 pts by factor 1, multiplied by 0.75
75.0 Glassene (id-nr 190)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 0 pts by factor 1, multiplied by 0.75
-12.5 briller (id-nr 188)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 0 pts by factor 1, multiplied by 0.5
-50.0 Myrseths (id-nr 187)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-50.0 kateteret (id-nr 180)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-62.5 lektor (id-nr 186)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
```

```
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-75.0 i (id-nr 197)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, -50 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
The proposed referent of anaphor DE is DEM
(wt-id reference): 198
(wt-features reference): ((de) (pron fl pers 3 akk @<p-utfyll @obj))
(wt-extra-features reference): #(9 0 0 190)
(wt-id anaphor): 201
(wt-ref anaphor): (176)
(wt-features anaphor): ((De) (pron fl pers 3 nom hoeflig @subj) (de)
                        (pron fl pers 3 nom @subj))
(wt-extra-features anaphor): #(2 0 0 190)
(WRONG!)

---------------------------------------------------------------------

Resolution of the sentence object completed
*********************************************************************

S-tokens: Det skulle vaere lett aa loese den naa naar jeg har gjennomgaatt et eksempel .

S-previous-1: - De kan ta oppgave 21 .

S-previous-2: Glassene var tykke , og det blinket i dem .


The considered anaphors in this sentence are:  "den".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor DEN has three resolution candidates.
The numbers in front are their point scores.

131.25 oppgave (id-nr 204)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 50 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
25.0 i (id-nr 197)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, -50 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-50.0 Glassene (id-nr 190)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor DEN is OPPGAVE
(wt-id reference): 204
(wt-features reference): ((oppgave) (subst appell mask ub ent @i-obj)
                          (oppgave) (subst appell fem ub ent @i-obj))
(wt-extra-features reference): #(5 0 0 -1)
(wt-id anaphor): 213
```

```
(wt-ref anaphor): (204 205)
(wt-features anaphor): ((den) (pron mask fem ent pers 3 @obj))
(wt-extra-features anaphor): #(7 0 0 204)
(CORRECT!)

------------------------------------------------------------------

Resolution of the sentence object completed
**************************************************************************

 Johan Ole reiste seg og gikk til tavlen .
There are no anaphors in this sentence.
**********************************************************
S-tokens: Han forsoekte aa huske noen av de ordene som hadde trengt seg inn til ham i hans
 dvaletilstand , men det var umulig aa faa noen sammenheng i dem .

S-previous-1:  Johan Ole reiste seg og gikk til tavlen .

S-previous-2: Det skulle vaere lett aa loese den naa naar jeg har gjennomgaatt et eksempel .


The considered anaphors in this sentence are:  "Han" "ham" "dem".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has three resolution candidates.
The numbers in front are their point scores.

150.0 Johan Ole (id-nr 223)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
-37.5 tavlen (id-nr 230)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-37.5 eksempel (id-nr 220)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
__________________________________________________________________
The proposed referent of anaphor HAN is JOHAN OLE
(wt-id reference): 223
(wt-features reference): ((Johan Ole) (subst prop mask @subj &person &org rx))
(wt-extra-features reference): #(2 0 3 -1)
(wt-id anaphor): 232
(wt-ref anaphor): (223)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(1 0 3 223)
(CORRECT!)

------------------------------------------------------------------

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
```

```
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAM has five resolution candidates.
The numbers in front are their point scores.

275 Han (id-nr 232)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
112.5 Johan Ole (id-nr 223)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
50 ordene (id-nr 239)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
0.0 tavlen (id-nr 230)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-37.5 eksempel (id-nr 220)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAM is HAN
(wt-id reference): 232
(wt-features reference): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features reference): #(1 0 3 223)
(wt-id anaphor): 246
(wt-ref anaphor): (232)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 akk @<p-utfyll))
(wt-extra-features anaphor): #(15 0 3 232)
(CORRECT!)

---------------------------------------------------------------------

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor DEM has seven resolution candidates.
The numbers in front are their point scores.

150 ordene (id-nr 239)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 0 pts by factor 1, multiplied by 1
25 sammenheng (id-nr 258)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
```

```
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
25 dvaletilstand (id-nr 249)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
0.0 tavlen (id-nr 230)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-37.5 Johan Ole (id-nr 223)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
-37.5 eksempel (id-nr 220)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-50 i (id-nr 259)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, -50 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, -100 pts by factor 1, multiplied by 1
_____________________________________________________________________
The proposed referent of anaphor DEM is ORDENE
(wt-id reference): 239
(wt-features reference): ((ord) (subst appell noeyt be fl @<p-utfyll))
(wt-extra-features reference): #(8 0 0 -1)
(wt-id anaphor): 260
(wt-ref anaphor): (239)
(wt-features anaphor): ((de) (pron fl pers 3 akk @<p-utfyll))
(wt-extra-features anaphor): #(29 0 0 239)
(CORRECT!)

---------------------------------------------------------------------

Resolution of the sentence object completed
*********************************************************************
(...)
```

B.2.4 Full analysis

The output of the `fullan` procedure is the most detailed level of presenting results. It shows the same analysis as the output of the `partan` preocedure, but also shows word-objects.

```
CL-USER(8): (fullan 11)
Johan Oles matematikklaerer holdt paa aa gjennomgaa en oppgave .
There are no anaphors in this sentence.
***********************************************************
Token: <Johan Oles>
ID-number: 0
Reference ID: NIL
Lemma: (Johan Oles)
Features: ((Johan Oles) (subst prop mask gen @det> &person
                         &sted &org &verk &hend &annet rx))
Previous: NIL
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: #(1 0 3 -1)
```

```
                Sentence possition: 1
                The word's lemma appearing in the title: No
                The word is denoting: human being of male sex

Token: <matematikklaerer>
ID-number: 2
Reference ID: NIL
Lemma: (matematikklaerer)
Features: ((matematikklaerer) (subst appell mask ub ent @subj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: #(2 0 0 -1)
                Sentence possition: 2
                The word's lemma appearing in the title: No
                The word is denoting: not a human being

Token: <holdt>
ID-number: -1
Reference ID: NIL
Lemma: (holde)
Features: ((holde) (verb pret tr1 tr11 pa1 a9 pr7 pr6 @fv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <paa>
ID-number: -1
Reference ID: NIL
Lemma: (paa)
Features: ((paa) (prep @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <aa>
ID-number: -1
Reference ID: NIL
Lemma: (aa)
Features: ((aa) (inf-merke @<p-utfyll))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <gjennomgaa>
ID-number: -1
Reference ID: NIL
Lemma: (gjennomgaa)
Features: ((gjennomgaa) (verb inf tr1 tr2 @iv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL
```

```
Token: <en>
ID-number: -1
Reference ID: NIL
Lemma: (en)
Features: ((en) (det mask ent kvant @det>))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <oppgave>
ID-number: 8
Reference ID: NIL
Lemma: (oppgave)
Features: ((oppgave) (subst appell mask ub ent @obj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: #(8 0 0 -1)
                 Sentence possition: 8
                 The word's lemma appearing in the title: No
                 The word is denoting: not a human being

Token: <.>
ID-number: -1
Reference ID: NIL
Lemma: ($.)
Features: (($.) (clb <<< <punkt>))
Previous: #<WORD-TOKEN>
Next: NIL
Antecedent: NIL
Points: NIL
Extra features: NIL


Gutten burde altsaa tenke paa matematikk .
There are no anaphors in this sentence.
************************************************************
Token: <Gutten>
ID-number: 10
Reference ID: NIL
Lemma: (gutt)
Features: ((gutt) (subst appell mask be ent @subj))
Previous: NIL
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: #(1 0 3 -1)
                 Sentence possition: 1
                 The word's lemma appearing in the title: No
                 The word is denoting: human being of male sex

Token: <burde>
ID-number: -1
Reference ID: NIL
Lemma: (burde)
Features: ((burde) (verb pret <aux1/infinitiv> @fv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
```

```
Points: NIL
Extra features: NIL

Token: <altsaa>
ID-number: -1
Reference ID: NIL
Lemma: (altsaa)
Features: ((altsaa) (adv @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <tenke>
ID-number: -1
Reference ID: NIL
Lemma: (tenke)
Features: ((tenke) (verb inf tr1 i1 pa1 tr2 tr12 tr3 rl8 rl7 @iv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <paa>
ID-number: -1
Reference ID: NIL
Lemma: (paa)
Features: ((paa) (prep @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <matematikk>
ID-number: 15
Reference ID: NIL
Lemma: (matematikk)
Features: ((matematikk) (subst appell mask ub ent @<p-utfyll))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: #(6 0 0 -1)
                 Sentence possition: 6
                 The word's lemma appearing in the title: No
                 The word is denoting: not a human being

Token: <.>
ID-number: -1
Reference ID: NIL
Lemma: ($.)
Features: (($.) (clb <<< <punkt>))
Previous: #<WORD-TOKEN>
Next: NIL
Antecedent: NIL
Points: NIL
Extra features: NIL
```

```
S-tokens: Men det gjorde han ikke .

S-previous-1: Gutten burde altsaa tenke paa matematikk .

S-previous-2: Johan Oles matematikklaerer holdt paa aa gjennomgaa en oppgave .


The considered anaphors in this sentence are:  "han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has five resolution candidates.
The numbers in front are their point scores.

150.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
50.0 Johan Oles (id-nr 0)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-37.5 oppgave (id-nr 8)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-37.5 matematikklaerer (id-nr 2)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
-56.25 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, -100 pts by factor 1, multiplied by 0.75
_____________________________________________________________________
The proposed referent of anaphor HAN is GUTTEN
(wt-id reference): 10
(wt-features reference): ((gutt) (subst appell mask be ent @subj))
(wt-extra-features reference): #(1 0 3 -1)
(wt-id anaphor): 20
(wt-ref anaphor): (10)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(4 0 3 10)
(CORRECT!)

---------------------------------------------------------------------

Resolution of the sentence object completed
*********************************************************************

Token: <Men>
ID-number: -1
Reference ID: NIL
Lemma: (men)
Features: ((men) (konj clb @kon))
```

```
Previous: NIL
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <det>
ID-number: 18
Reference ID: (13 14 15)
Lemma: (det)
Features: ((det) (pron noeyt ent pers 3 @obj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <gjorde>
ID-number: -1
Reference ID: NIL
Lemma: (gjoere)
Features: ((gjoere) (verb pret tr1 rl9 pr3 @fv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <han>
ID-number: 20
Reference ID: (10)
Lemma: (han)
Features: ((han) (pron mask ent pers hum 3 nom @subj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: #<WORD-TOKEN>
     Antecedent's word-token:  <GUTTEN>
     Antecedent's ref-number: NIL
Points: ((20 11 0) (20 10 0) (20 9 50) (20 7 0) (20 6 0) (20 5 0) (20 2 100) (20 1 100))
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 20.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 20.
        Awarded 50 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 20.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 20.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 20.
        Awarded 0 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 20.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 20.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 20.
Extra features: #(4 0 3 10)
                Sentence possition: 4
                The word's lemma appearing in the title: No
                The word is denoting: human being of male sex

                The pronoun is bound to the
                    antecedent with id-number 10
```

```
Token: <ikke>
ID-number: -1
Reference ID: NIL
Lemma: (ikke)
Features: ((ikke) (adv @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <.>
ID-number: -1
Reference ID: NIL
Lemma: ($.)
Features: (($.) (clb <<< <punkt>))
Previous: #<WORD-TOKEN>
Next: NIL
Antecedent: NIL
Points: NIL
Extra features: NIL


S-tokens: Hvis noen hadde spurt ham hva han tenkte paa , ville han ikke kunnet svare .

S-previous-1: Men det gjorde han ikke .

S-previous-2: Gutten burde altsaa tenke paa matematikk .


The considered anaphors in this sentence are:  "ham" "han" "han".

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAM has three resolution candidates.
The numbers in front are their point scores.

187.5 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
50.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAM is HAN
(wt-id reference): 20
(wt-features reference): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features reference): #(4 0 3 10)
```

```
(wt-id anaphor): 27
(wt-ref anaphor): (20)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 akk @obj))
(wt-extra-features anaphor): #(5 0 3 20)
(CORRECT!)

-----------------------------------------------------------------

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has four resolution candidates.
The numbers in front are their point scores.

325 ham (id-nr 27)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
225.0 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
75.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAN is HAM
(wt-id reference): 27
(wt-features reference): ((han) (pron mask ent pers hum 3 akk @obj))
(wt-extra-features reference): #(5 0 3 20)
(wt-id anaphor): 29
(wt-ref anaphor): (27)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(7 0 3 27)
(CORRECT!)

-----------------------------------------------------------------

Factor 1 (GENDER/NUMBER/PERSON) applied.
Factor 2 (REFERENCE PROXIMITY) applied.
Factor 5 (DIRECT OBJECT) applied.
Factor 6 (INDIRECT OBJECT NP) applied.
Factor 7 (ADVERBIAL NP) applied.
Factor 9 (SYNTACTIC PARALLELISM) applied.
Factor 10 (SECTION HEADING PREFERENCE) applied.
Factor 11  (INDEFINITNESS) applied.
Factor 3 (BOOST PRONOUN) applied.
The anaphor HAN has five resolution candidates.
The numbers in front are their point scores.

325 han (id-nr 29)
```

```
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
325 ham (id-nr 27)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 50 pts by factor 5,
 100 pts by factor 2, 100 pts by factor 1, multiplied by 1
225.0 han (id-nr 20)
     Points: 75 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 50 pts by factor 2, 100 pts by factor 1, multiplied by 0.75
75.0 Gutten (id-nr 10)
     Points: 0 pts by factor 3, 0 pts by factor 11, 0 pts by factor 10,
 50 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, 100 pts by factor 1, multiplied by 0.5
-62.5 matematikk (id-nr 15)
     Points: 0 pts by factor 3, -25 pts by factor 11, 0 pts by factor 10,
 0 pts by factor 9, 0 pts by factor 7, 0 pts by factor 6, 0 pts by factor 5,
 0 pts by factor 2, -100 pts by factor 1, multiplied by 0.5
_____________________________________________________________________
The proposed referent of anaphor HAN is HAN
(wt-id reference): 29
(wt-features reference): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features reference): #(7 0 3 27)
(wt-id anaphor): 34
(wt-ref anaphor): (29)
(wt-features anaphor): ((han) (pron mask ent pers hum 3 nom @subj))
(wt-extra-features anaphor): #(12 0 3 29)
(CORRECT!)

---------------------------------------------------------------------

Resolution of the sentence object completed
*********************************************************************

Token: <Hvis>
ID-number: -1
Reference ID: NIL
Lemma: (hvis)
Features: ((hvis) (sbu @adv))
Previous: NIL
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <noen>
ID-number: 24
Reference ID: (-1)
Lemma: (noen)
Features: ((noen) (pron mask fem ent pers 3 @subj) (noen) (pron fl pers 3 @subj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <hadde>
ID-number: -1
Reference ID: NIL
Lemma: (ha)
Features: ((ha) (verb pret pa1 d5 a6 tr12 rl6 pa3 tr6 d6/til <aux1/perf_part> pa6 @fv))
```

```
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <spurt>
ID-number: -1
Reference ID: NIL
Lemma: (spoerre)
Features: ((spoerre) (verb perf-part tr1 i1 tr11 pa1 d5 rl9 pa2 @iv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <ham>
ID-number: 27
Reference ID: (20)
Lemma: (han)
Features: ((han) (pron mask ent pers hum 3 akk @obj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: #<WORD-TOKEN>
     Antecedent's word-token:  <HAN>
     Antecedent's ref-number: (10)
Points: ((34 3 75) (34 11 0) (34 10 0) (34 9 0) (34 7 0) (34 6 0)
         (34 5 50) (34 2 100) (34 1 100) (29 3 75) (29 11 0)
         (29 10 0) (29 9 0) (29 7 0) (29 6 0) (29 5 50) (29 2 100)
         (29 1 100) (27 11 0) (27 10 0) (27 9 50) (27 7 0)
         (27 6 0) (27 5 50) (27 2 100) (27 1 100))
        Awarded 75 points by BOOST PRONOUN factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 34.
        Awarded 50 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 34.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 34.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 34.
        Awarded 75 points by BOOST PRONOUN factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 29.
```

```
        Awarded 50 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 29.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 29.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 27.
        Awarded 50 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 27.
        Awarded 50 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 27.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 27.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 27.
Extra features: #(5 0 3 20)
                Sentence possition: 5
                The word's lemma appearing in the title: No
                The word is denoting: human being of male sex

                The pronoun is bound to the
                    antecedent with id-number 20

Token: <hva>
ID-number: 28
Reference ID: (-1)
Lemma: (hva)
Features: ((hva) (pron sp @obj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <han>
ID-number: 29
Reference ID: (27)
Lemma: (han)
Features: ((han) (pron mask ent pers hum 3 nom @subj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: #<WORD-TOKEN>
     Antecedent's word-token:  <HAM>
     Antecedent's ref-number: (20)
Points: ((34 3 75) (34 11 0) (34 10 0) (34 9 50) (34 7 0) (34 6 0)
         (34 5 0) (34 2 100) (34 1 100) (29 11 0) (29 10 0)
         (29 9 50) (29 7 0) (29 6 0) (29 5 0) (29 2 100) (29 1 100)
         (27 11 0) (27 10 0) (27 9 0) (27 7 0) (27 6 0)
         (27 5 0) (27 2 100) (27 1 100))
        Awarded 75 points by BOOST PRONOUN factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 34.
```

```
 Awarded 50 points by SYNTACTIC PARALLELISM factor
while resolving the anaphor with id-number 34.
 Awarded 0 points by ADVERBIAL NP factor
while resolving the anaphor with id-number 34.
 Awarded 0 points by INDIRECT OBJECT NP factor
while resolving the anaphor with id-number 34.
 Awarded 0 points by DIRECT OBJECT factor
while resolving the anaphor with id-number 34.
 Awarded 100 points by REFERENCE PROXIMITY factor
while resolving the anaphor with id-number 34.
 Awarded 100 points by GENDER/NUMBER/PERSON factor
while resolving the anaphor with id-number 34.
 Awarded 0 points by INDEFINITNESS factor
while resolving the anaphor with id-number 29.
 Awarded 0 points by SECTION HEADING PREFERENCE factor
while resolving the anaphor with id-number 29.
 Awarded 50 points by SYNTACTIC PARALLELISM factor
while resolving the anaphor with id-number 29.
 Awarded 0 points by ADVERBIAL NP factor
while resolving the anaphor with id-number 29.
 Awarded 0 points by INDIRECT OBJECT NP factor
while resolving the anaphor with id-number 29.
 Awarded 0 points by DIRECT OBJECT factor
while resolving the anaphor with id-number 29.
 Awarded 100 points by REFERENCE PROXIMITY factor
while resolving the anaphor with id-number 29.
 Awarded 100 points by GENDER/NUMBER/PERSON factor
while resolving the anaphor with id-number 29.
 Awarded 0 points by INDEFINITNESS factor
while resolving the anaphor with id-number 27.
 Awarded 0 points by SECTION HEADING PREFERENCE factor
while resolving the anaphor with id-number 27.
 Awarded 0 points by SYNTACTIC PARALLELISM factor
while resolving the anaphor with id-number 27.
 Awarded 0 points by ADVERBIAL NP factor
while resolving the anaphor with id-number 27.
 Awarded 0 points by INDIRECT OBJECT NP factor
while resolving the anaphor with id-number 27.
 Awarded 0 points by DIRECT OBJECT factor
while resolving the anaphor with id-number 27.
 Awarded 100 points by REFERENCE PROXIMITY factor
while resolving the anaphor with id-number 27.
 Awarded 100 points by GENDER/NUMBER/PERSON factor
while resolving the anaphor with id-number 27.
Extra features: #(7 0 3 27)
                Sentence possition: 7
                The word's lemma appearing in the title: No
                The word is denoting: human being of male sex

                The pronoun is bound to the
                    antecedent with id-number 27

Token: <tenkte>
ID-number: -1
Reference ID: NIL
Lemma: (tenke)
Features: ((tenke) (verb pret tr1 i1 pa1 tr2 tr12 tr3 rl8 rl7 @fv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL
```

```
Token: <paa>
ID-number: -1
Reference ID: NIL
Lemma: (paa)
Features: ((paa) (prep @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <,>
ID-number: -1
Reference ID: NIL
Lemma: ($,)
Features: (($,) (clb <komma> @s-gr))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <ville>
ID-number: -1
Reference ID: NIL
Lemma: (ville)
Features: ((ville) (verb pret tr1 tr2 d1 <aux1/infinitiv> @fv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <han>
ID-number: 34
Reference ID: (29)
Lemma: (han)
Features: ((han) (pron mask ent pers hum 3 nom @subj))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: #<WORD-TOKEN>
     Antecedent's word-token:  <HAN>
     Antecedent's ref-number: (27)
Points: ((34 11 0) (34 10 0) (34 9 50) (34 7 0) (34 6 0)
         (34 5 0) (34 2 100) (34 1 100) (29 11 0) (29 10 0)
         (29 9 50) (29 7 0) (29 6 0) (29 5 0) (29 2 100)
         (29 1 100) (27 11 0) (27 10 0) (27 9 0) (27 7 0)
         (27 6 0) (27 5 0) (27 2 100) (27 1 100))
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 34.
        Awarded 50 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 34.
        Awarded 100 points by REFERENCE PROXIMITY factor
```

```
       while resolving the anaphor with id-number 34.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 34.
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 29.
        Awarded 50 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 29.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 29.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 29.
        Awarded 0 points by INDEFINITNESS factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by SECTION HEADING PREFERENCE factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by SYNTACTIC PARALLELISM factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by ADVERBIAL NP factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by INDIRECT OBJECT NP factor
       while resolving the anaphor with id-number 27.
        Awarded 0 points by DIRECT OBJECT factor
       while resolving the anaphor with id-number 27.
        Awarded 100 points by REFERENCE PROXIMITY factor
       while resolving the anaphor with id-number 27.
        Awarded 100 points by GENDER/NUMBER/PERSON factor
       while resolving the anaphor with id-number 27.
Extra features: #(12 0 3 29)
                Sentence possition: 12
                The word's lemma appearing in the title: No
                The word is denoting: human being of male sex

                The pronoun is bound to the
                    antecedent with id-number 29

Token: <ikke>
ID-number: -1
Reference ID: NIL
Lemma: (ikke)
Features: ((ikke) (adv @adv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <kunnet>
ID-number: -1
Reference ID: NIL
Lemma: (kunne)
Features: ((kunne) (verb perf-part tr1 tr3 <aux1/infinitiv> @iv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
```

```
Points: NIL
Extra features: NIL

Token: <svare>
ID-number: -1
Reference ID: NIL
Lemma: (svare)
Features: ((svare) (verb inf tr1 i1 tr11 rl5 tr2 d1 d2 @iv))
Previous: #<WORD-TOKEN>
Next: #<WORD-TOKEN>
Antecedent: NIL
Points: NIL
Extra features: NIL

Token: <.>
ID-number: -1
Reference ID: NIL
Lemma: ($.)
Features: (($.) (clb <<< <punkt>))
Previous: #<WORD-TOKEN>
Next: NIL
Antecedent: NIL
Points: NIL
Extra features: NIL
(...)
```

References

Alshawi, H. (1987). *Memory and context for language interpretation*. Studies in natural language processing. Cambridge: Cambridge University Press.

Aone, C. and S. W. Bennett (1995). Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pp. 122–129. Association for Computational Linguistics.

Baldwin, B. (1995). *CogNIAC: A Discourse Processing Engine*. Ph. D. thesis, University of Pennsylvania Department of Computer and Information Sciences.

Brennan, S., M. W. Friedman, and C. J. Pollard (1987). A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association of Computational Linguistics*.

Carbonell, J. G. and R. D. Brown (1988). Anaphora resolution: A multy-strategy approach. In *Proceedings of 12th International Conference on Computational Linguistics (COLING'88)*, Budapest, Hungary, pp. 96–101.

Carter, D. (1987). *Interpreting anaphors in natural language texts*. Chichester: Ellis Horwood ltd.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Dagan, I. and A. Itai (1990). Automatic processing of large corpora for the resolution of anaphora references. In *Proceedings of the 13th conference on Computational linguistics*, pp. 330–332. Association for Computational Linguistics.

Dagan, I., J. S. Justeson, S. Lappin, H. J. Leass, and A. Ribak (1995). Syntax and lexical statistics in anaphora resolution. *Applied Artificial Intelligence 9*(6), 633–644.

Denber, M. (1998). Automatic resolution of anaphora in English. Technical report, Eastman Kodak Co., Imaging Science Division.

Faarlund, J. T., S. Lie, and K. I. Vannebo (1997). *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Firbas, J. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge University Press.

Fraurud, K. (1992). *Processing noun phrases in natural discourse*. Ph. D. thesis, Department of Linguistics, Stockholm University.

Ge, N., J. Hale, and E. Charniak (1998). A statistical approach to anaphora resolution. In *Proceedings of the Workshop on Very Large Corpora*, Montreal, Canada, pp. 161–170.

Grosz, B. J., A. K. Joshi, and S. Weinstein (1995). Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics 21*(2), 203–226.

Gundel, J. K. (2002). Information structure and the use of cleft sentences in English and Norwegian. In H. Hasselgård, S. K. A. Johansson, B. Behrens, and C. Fabricius-Hansen (Eds.), *Information Structure in a Cross-linguistic Perspective*, Volume 39 of *Language and Computers*, pp. 113–129. Amsterdam: Rodopi.

Halliday, M. A. K. and R. Hasan (1976). *Cohesion in English*. English language series. London: Longman.

Hassel, M. (2000). Pronominal resolution in automatic text summarisation. Master's thesis, Department of Computer and System Sciences, Stockholm University.

Hirst, G. (1981). Discourse-oriented anaphora resolution in natural language understanding: A review. *American Journal of Computational Linguistics 7*(2), 85–98.

Hobbs, J. R. (1978). Resolving pronoun references. *Lingua 44*, 311–338.

Johansson, M. (2001). Clefts in contrast. A contrastive study of English and Swedish texts and translations. *Linguistics 39*, 547–582.

Jónsdóttir, A. B. (2003, May). ARNER, what kind of name is that? - an automatic rule-based name entity recognizer for Norwegian. Master's thesis, University of Oslo, Department of Linguistics.

Kamp, H. (1981). A theory of truth and semantic representation: Part 1. In J. Groenendijk, T. Janssen, and M. Stokhof (Eds.), *Formal Methods in the Study of Language*, pp. 277 – 322. Mathematisch Centrum.

Kennedy, C. and B. Boguraev (1996). Anaphora for everyone: Pronominal anaphora resolution without a parser. In *The Proceedings of the 16th Inter-*

national Conference on Computational Linguistics (COLING'96), Copenhagen, Denmark, pp. 113–118.

Lappin, S. and H. J. Leass (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics 20*(4), 535–561.

Lappin, S. and M. McCord (1990). A syntactic filter on pronominal anaphora in slot grammar. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pp. 135–142.

Lødrup, H. (1999). Linking and optimality in Norwegian presentational focus construction. *Nordic journal of linguistics 22*, 205–230.

McCord, M. C. (1990). Slot grammar: A system for simpler construction of practical natural language grammars. In R. Studer (Ed.), *Proceedings of the International Symposium on Natural Language and Logic*, pp. 118–145. Springer-Verlag.

Mikheev, A., M. Moens, and C. Grover (1999). Named entity recognition without gazetteers. In *Proceedings if the 9th European Chapter of the Association of Computational Linguists (EACL)*, Bergen, Norway, pp. 1–8.

Mitkov, R. (1997). Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches. In *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, Madrid, Spain, pp. 14–21.

Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *COLING-ACL*, pp. 869–875.

Mitkov, R. (2001). Outstanding issues in anaphora resolution. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*, pp. 110–125. Springer.

Mitkov, R. (2002). *Anaphora Resolution*. Pearson Education.

Mitkov, R., B. Boguraev, and S. Lappin (2001). Introduction to the special issue on computational anaphora resolution. *Computational Linguistics 27*(4), 473–477.

Orasan, C., R. Evans, and R. Mitkov (2000). Enhancing preference-based anaphora resolution with genetic algorithms. In *Natural Language Processing*, pp. 185–195.

Øvrelid, L. (2003, November). Subject or object? Syntactic disambiguation in Norwegian, stochastic optimality theory and application in an automatic system. Master's thesis, University of Oslo, Department of Linguistics.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.

Rich, E. and S. LuperFoy (1988). An architecture for anaphora resolution. In *ACL Proceedings, Second Conference on Applied Natural Language Processing*, pp. 18–24.

Strube, M. and U. Hahn (1999). Functional centering – Grounding referential coherence in information structure. *Computational Linguistics 25*(3), 309–344.

Stuckardt, R. (2001). Design and enhanced evaluation of a robust anaphor resolution algorithm. *Computational Linguistics 27*(4), 479–506.

Vieira, R. and M. Poesio (2000). An empirically-based system for processing definite descriptions.

Winograd, T. (1972). *Understanding Natural Language*. Orlando, FL, USA: Academic Press, Inc.